

Continuous Compliance Home

Continuous Compliance

Exported on 08/02/2023

Table of Contents

Welcome to the Delphix Continuous Compliance documentation!	5
Release notes	6
New features	7
Fixed issues.....	27
Known issues.....	65
Deprecated and end-of-life features	72
Licenses and notices.....	75
Getting started	76
Introduction to Delphix Masking.....	77
Data source support.....	80
Installation	93
Naming requirements.....	132
Users and roles.....	136
Best practices for defining masking roles.....	156
Audit logs.....	159
Kerberos configuration.....	161
Password vault configuration	174
DB2 connector license installation	178
Continuous Compliance Engine icon reference	180
Delphix masking terminology	181
Changing the IP address of the Delphix Engine.....	187
Stopping and starting the containerized Continuous Compliance Engine	189
Stopping, starting, and restarting the continuous compliance engine	191
Upgrading the Delphix Continuous Compliance Engine.....	193
Utilization	194
Preparing data	196
Database user permissions for executing masking and profiling job	197
Preparing Oracle database for profiling/masking.....	198
Preparing SQL server database for profiling and masking	201
Preparing Sybase database for profiling and maskin	203
Connecting data	206
Managing environments	207

Managing remote mounts for VM continuous compliance engines	216
Managing remote mounts for containerized masking	222
Managing SSL/TLS over JDBC for containerized masking.....	226
Managing connectors	229
Managing extended connectors	239
Managing rule sets	247
Managing file formats	258
Managing inventories	262
Managing record types	274
Masking whole file.....	277
JSON file masking	280
Identifying sensitive data	288
Discovering your sensitive data	289
Out of the box profiling settings.....	291
ASDD standard profile set.....	292
Standard profile set expressions.....	318
Legacy profile set expressions.....	325
Configuring profile sets.....	335
Managing domains.....	339
Managing classifiers.....	342
Managing expressions	347
ASDD profile set import and export	351
Creating a profiling job	352
Running a profiling job	355
Reporting profiling results.....	356
ASDD features and support	364
Securing sensitive data.....	366
Algorithms	367
Builtin Driver Supports	491
Creating masking jobs	498
Managing jobs	503
Monitoring masking job.....	505
Masking job wizard	512
Running stopping jobs.....	520

Masked provisioning	521
Configuring virtualization service for masked provisioning.....	522
Provision masked VDBs.....	523
Authoring extensible plugins	530
Introduction (Authoring extensible plugins).....	531
General plugin structure.....	533
Setting up your development environment.....	540
Algorithms (Authoring extensible plugins).....	542
Driver supports.....	590
Managing plugins using the API client.....	608
Installing a plugin onto the Delphix masking engine.....	609
Secure plugin deployment.....	610
Terminology.....	612
Delphix masking APIs	613
Masking client.....	614
API examples.....	717
Managing multiple engines for masking	741
Introduction (Managing multiple engines for masking).....	742
Sync concepts.....	745
Sync endpoints.....	754
Key management.....	758
Algorithm syncability.....	760
User workflow examples.....	763
Change log.....	777

Welcome to the Delphix Continuous Compliance documentation!

Here you will find all the information you will need to use the Delphix Continuous Compliance Engine. Learn how to deploy our application, use our features, or tune it for optimal performance. We have organized this content into several categories that you can browse via the navigation bar.

The following is a list of the PDF copy of all the Continuous Compliance documentation versions:

Release notes

This section covers the following topics:

- [New features](#)
- [Fixed issues](#)
- [Known issues](#)
- [Deprecated and end-of-life features](#)
- [Licenses and notices](#)

New features

Release 13.0.0.0

- **Password Vault Support for Microsoft SQL Server**
In this release, the SQL Server connector joins the growing list of connectors that support credential management systems.
- **SQL Server**
SQL Server 2022 is now supported.
- **Microsoft Azure SQL Data Warehouse**
Azure SQL Data Warehouse is now supported.
- **Updated User Experience**
The Roles page has been updated, with a number of handy templates now provided as well.

Release 12.0.0.0

- **ASDD extended drivers**
Extended drivers allow you to provide your driver with additional data sources. Automated Sensitive Data Discovery, which aims to identify common sensitive and personally identifiable data elements, now works with these sources.
- **Password vault support for Oracle**
You can now leverage supported password vaults to manage credentials for Oracle databases.

Release 11.0.0.0

- **IBAN discovery and masking**
Supports automatically discovering and masking International Bank Account Numbers.
- **New US driver's license classifier**
Supports automatically discovering and masking US Driver's Licenses.
- **Oracle source sizing**
Measures the size of connected Oracle databases. This information will be available via a new utilization PDF report, as well as through the GET /billing-usage API.
- **New features for the Postgres connector**
The connector will now provide more automated control of indexes, constraints, and triggers. This will automatically upgrade and allow you to choose these features.
- **User-controlled automated sensitive data discovery (ASDD) set upgrade**
With the 11.0 release, Delphix will begin shipping the standard ASDD discovery set independent of Continuous Compliance. You can use a new API endpoint to control when to adopt new ASDD discovery sets. This will allow you to upgrade to newer releases of the software without impacting existing workflows, as well as easily creating and distributing your own classifier sets.

Release 10.0.0.0

- **Password vault support for PostgreSQL connections**
This release adds HashiCorp Vault and CyberArk support for storing and accessing secrets, keys, and certificates necessary for PostgreSQL operations.
- **More automated sensitive data discovery classifiers**
Classifiers improve our ability to automatically discover sensitive information and accurately recommend masking algorithms. This update adds classifiers for identifying Credit Cards, Addresses, Telephone

numbers, US Passports, US SSNs, and Bank accounts. With ASDD Classifiers, organizations can easily mask new data sources for downstream teams to start working with compliant data quickly.

- **Classifiers user interface**

This update adds a new user interface for the Automated Sensitive Data Discovery classifiers introduced in 9. Previously, classifier definitions were only accessible via the API.

- **Enhanced document type support for JSON and XML**

This update expands our support for the JSON and XML Document types. Applying the multi-column algorithms to JSON and XML structures is now possible. Further, it is now possible to mask JSON and XML structures stored in BLOBs.

- **CData license upload**

Our partnership with CData allows customers to purchase and use CData JDBC drivers from Delphix to connect to additional data sources with Continuous Compliance. This release enables users to install the required license files on their engines. Customers download the CData drivers directly from CData but download the corresponding license files for these drivers from Delphix. Consult your account representative for purchasing and using a specific CData driver.

Release 9.0.0.0

- **New Automated Sensitive Data Discovery (ASDD) profiler implementation**

The new ASDD profiler has been introduced with a number of [new features](#), including new data-level profiling logic and improved database sampling capabilities. A new profile set, [ASDD Standard](#) has been added that leverages the new implementation.

- **UI Updates: Audit and Classifiers**

Audit: This page allows for a quick review of historical user activity with global audit logs, to help ensure compliance. Updated for speed and ease-of-use, with significantly better filtering, searching, and performance.

Classifiers: This section of the Settings page contains a list of Classifiers with relevant information and actions. The Profiler can be customized for domain and data level (name, column metadata, data value) using standard regex expressions. Improved performance, data level inspection, and accuracy for matching data types (using data classifiers).

Release 8.0.0.0

- **Improved JSON document store performance**

Upon upgrade, JSON Document Store masking is significantly faster. For more information, visit [Document Store Type Masking](#).

- This release contains bug fixes for the Continuous Compliance Engine.

Release 7.0.0.0

- **JSON and XML: tokenization/re-identification and chained algorithms**

Easily leverage tokenization/re-identification algorithms and chained algorithms with JSON and XML data. Data tokenization/re-identification provides reversible data anonymization to protect data in non-prod environments. Chain algorithms enable complex multistep algorithms to be run on separate values, such as Full Name algorithms.

Release 6.0.17

- **JSON; XML field masking**

This release introduces the ability to mask JSON; XML objects nested in string/text fields, allowing Continuous Compliance to be maintained for semi-structured JSON & XML data in non-production

environments. With this, the Continuous Compliance library can be leveraged, or customized algorithms can be created to meet required data schemas.

- **Microsoft Intelligent Data Platform Integration (Azure Data Factory; Azure Synapse Pipelines)**
Accelerate data compliance using Microsoft Intelligent Data Platform's ETL tools with Delphix Continuous Compliance. This allows users to quickly mask data while moving between 100+ Azure Synapse Analytics and Azure Data Factory connections. Quickly identify sensitive data in ETL pipelines and mask using Delphix algorithms.
- **Containerized masking**
This feature allows users to efficiently spin up and tear down Continuous Compliance containers. Easily orchestrate scaling out Continuous Compliance using a container orchestrator, allowing for quick parallelized masking jobs in the self-managed container cluster to multiple instances.
- **Hyperscale Compliance masking job sync**
Introducing fast migration for masking jobs from existing Continuous Compliance Engines to the Hyperscale Compliance Orchestrator. This accelerates the masking of massive Oracle databases for compliance and greatly improves masking speed for databases with billions of rows containing PII, PHI, or sensitive data fields.

Release 6.0.16.0

- **Strict content security policy**
This release adds a new application setting group that drives strict content security policy. Content Security Policy (CSP) is an added layer of security that helps to detect and mitigate certain types of attacks. Including (but not limited to) Cross Site Scripting (XSS), and data injection attacks.

Release 6.0.15.0

- **JSON file masking**
This update introduces support for JSON File Masking, a popular human-readable data exchange format. Teams can leverage Delphix's existing library of pre-built algorithms to mask sensitive information. For more information, see [JSON File Masking](#).
- **Segmented mapping algorithm V2**
This update enhances the Segmented Mapping Algorithm for improved performance, extensibility, security, and portability. It produces new masking results from the legacy algorithm. Segmented Mapping allows a user to create unique masked values by dividing data into segments that are masked piecewise. For more information, see [Segmented Mapping](#).
- **Numeric expression algorithm**
This new algorithm enables formula transformations to values, so that teams can run common math operators to scale or shift numbers.
- **New API endpoints**
This release introduced a new API Support Bundle Generation endpoint:

The new API endpoint is:

Group	Endpoints	Description
supportBundle	POST /support-bundle	Generates support bundle

For more information, see [API for generating support bundle](#).

Release 6.0.14.0

- **Certifications**

- SAP HANA 2.0 SP 05
- CockroachDB
- VMware ESXi 7.0 U3c
- **Data cleansing**
The Data Cleansing algorithm has been updated to standardize spellings, misspellings, and convert abbreviations. Algorithm based data cleansing eliminates slow and manual processes prior to masking the data. For more information, see [Data Cleansing](#).
- **Min Max**
The MinMax Number and MinMax Date algorithms have been updated for normalizing outlier data in a table column by masking numbers and dates that could give context of records based on values.
- **Continuous Compliance UI improvements**
 - Monitoring interface now details which job step is currently underway
 - Improved Continuous Compliance job management for intermediate steps
 - Redesigned Execution Monitor interface
 - For more information, see [Monitoring Masking Job](#).
- **“Optional” columns for multi-column algorithms**
The Multi-column algorithm now allows for some fields to be marked as optional for more run-time flexibility. Concurrent Continuous Compliance operations can now occur using multiple data columns in a single operation. For more information, see [Using Multi-Column Algorithms](#).
- **Automated sensitive data discovery**
A new default profile set is being implemented to improve the accuracy and speed of column level profiling. 40 new profile expressions utilizing type constraints are introduced for domain and algorithm assignment to help reduce false positives associated with data type mismatches during inspection. The default profiler set upgrade has no change or impact on existing users.
- **New API endpoints**
This release extends the list of API-endpoints by adding the following task-based progress monitoring endpoints.
The new API endpoints are :

Group	Endpoints	Description
monitoring	GET /monitor-task	get the status of Execution or async task
monitoring	GET /monitor-task/	get the status of Execution or async task by the id

Release 6.0.13.0

- **Certifications**
This release adds support for VMware ESX/ESXi 7.0 U3c and DB2 12.0 on z/OS.
- **New tokenization algorithm**
In this release, Delphix introduces a new Tokenization algorithm framework to replace the legacy Tokenization algorithm. This new Tokenization algorithm framework includes additional configuration options for increased security. For more information, see [Tokenization](#). Legacy Tokenization algorithm instances will remain in place and function the same until their planned EOL in version 6.0.15.0, migration to the new Tokenization algorithm is recommended.
- **Zip+4 algorithm**
A new version of the Zip+4 algorithm is now available that is used for full-length (nine-digit) zip codes. This new version is built upon the Masking Algorithm SDK and offers the same benefits as other new algorithms, including greater performance.

- **Improved masking monitoring**

This release improves usability and diagnosability of the Masking Engine by allowing users to search for past jobs and filter the results based on job type and status. For more information, see the **Search** section at [Monitoring Masking Job](#).

- **New API endpoints**

This release extends the list of API-endpoints by adding the following execution logs endpoints. These API endpoints will return file download ID that can be used to download execution logs under GET /file-downloads/{fileDownloadId}.

The new API endpoints are :

Group	Endpoints	Description
executions	GET /execution-logs	get all execution logs of all jobs.
executions	GET /executions/	get a particular execution log by using execution ID.
execution-component-log	GET /execution-component-log	get all the execution component logs of all jobs, execution ID is a mandatory parameter.
execution-component-log	GET /execution-component-log/	get a particular execution log by using componentId.

Release 6.0.12.0

- **New Microsoft SQL Server implementation to disable constraints/triggers and drop indexes**

In this release, Delphix adds default driver support for Microsoft SQL Server database masking options of Disable Constraints, Drop Indexes, and Disable Triggers as job tasks. These changes apply to masking, reidentification, and tokenization jobs where enabled.

For details on the usage and known limitations of the Microsoft SQL Server Disable Constraints, Drop Indexes, and Disable Triggers driver support tasks, see [Microsoft SQL Server Built-in Driver Support Plugin](#).

Upon engine upgrade, any existing jobs on built-in Microsoft SQL Server connectors where these options were selected will be upgraded to these enabled driver support plugin tasks.

- **Improved user experience diagnosability**

This release improves user experience by ensuring time is displayed in a consistent fashion. It prevents users from running parallel update threads against incompatible databases and provides per-job log information. The job monitoring view now displays the total time taken in the hours:minutes:seconds format.

- **Free text redaction**

This release updates the free text redaction algorithm to the new extensible Algorithm framework to improve performance and allow chaining of algorithm instances. For more information, see [Free Text Redaction](#).

- **Updated secure lookup instances**

This release updates the legacy Secure Lookup algorithms to the extensible Secure Lookup framework. Masked results will remain the same other than whitespace handling.

- **New API endpoint for define fields**

This release extends the list of API-endpoints by adding the following file-field-metadata endpoint to create field metadata for a file format. This field allows users to add a file field that you want to mask in a format. After the user uploads a format, all the fields from the uploaded file format are displayed at the inventory screen.

The new API endpoint is :

Group	Endpoints	Description
fileFieldMetadata	POST /file-field-metadata	Creates field metadata for a file format.

- **Masking whole file**
You can now configure the masking engine to mask the complete file and pass the content of that file as a single input to an algorithm. For more information, see [Masking Whole File](#).
- **Character mapping algorithm support for tokenization/reidentification jobs**
The character mapping algorithm can now be used for tokenization and reidentification jobs.

Release 6.0.11.0

- **Certifications**
This release adds support for Oracle database 21c.
- **General UI for extended algorithms**
In this release, Delphix continues to improve the experience of creating and using new extended algorithms. These algorithms may include configuration information stored in JSON format. The configurations are now editable via the UI. For more information, see [General UI for Extended Algorithms](#).
- **OAuth2 API support**
The Virtualization and Masking engine APIs are now accessible via OAuth2 tokens that improve Delphix's security offerings. For more information, see [Configuring OAuth2 Authentication for API Access](#).
- **New Oracle optimizations to disable constraints/triggers and drop indexes**
In this release, Delphix has re-implemented the Oracle database masking options of Disable Constraints, Drop Indexes, and Disable Triggers as job tasks, using the [Driver Support Plugin Framework](#), improving both functionality and performance. These optimizations apply to masking, reidentification, and tokenization jobs where these tasks are enabled.
For details on the optimizations, usage, and known limitations of the Oracle Disable Constraints, Drop Indexes, and Disable Triggers driver support tasks, see [Oracle Built-in Driver Support Plugin](#). Upon engine upgrade, any existing jobs on built-in Oracle connectors where these options were selected will be upgraded to these enabled driver support plugin tasks.
- **New export secure lookup values API**
This release extends the list of API-endpoints by adding a new API for exporting the values from a secure lookup algorithm instance.
The new API endpoint is:

Group	Endpoints	Description
algorithm	POST /algorithms/	Export lookup values form secure lookup algorithm.

For more information, see [Secure Lookup - Exporting Secure Lookup Values via API](#).

- **New copy ruleset API**
This release extends the list of API-endpoints by adding three new APIs for copying rulesets under databaseRuleset, fileRuleset, and mainframeDatasetRuleset.
The new API endpoints are :

Group	Endpoints	Description
databaseRuleset	PUT /database-rulesets/	Copy ruleset objects in the same database environment.
fileRuleset	PUT /file-rulesets/	Copy ruleset objects in the same file environment.
mainframeDatasetRuleset	PUT /mainframe-dataset-rulesets/	Copy ruleset objects in the same dataset environment.

- **New binary lookup algorithm**

This release introduces a new binary lookup algorithm framework in the masking extensibility SDK that supports advanced features such as algorithm chaining. Legacy binary lookup algorithm instances will be automatically and seamlessly migrated to the new binary lookup framework when you upgrade the masking engine. For more information, see [Binary Lookup](#).

- **UI/UX enhancements**

This release introduces substantial improvements to the user interface that gives a new look and feel to the masking engine.

Release 6.0.10.0

- **Masking Salesforce data**

There has been an increasing demand for an easy way to manage and utilize the highly sensitive data stored in Salesforce. With this new Select Connector offering, sensitive data discovery and masking algorithm assignment is automatically handled for the Salesforce default schema; this is not only unique in the market, but also the first time Delphix is delivering this solution as an addition to its product suite. This is the top compliance solution for Salesforce on the market and provides a dramatically simpler deployment option to manage and secure this business-critical data. For more information, see [Application Solutions documentation](#).

- **New mapping algorithm**

A more powerful and faster mapping algorithm is now available. This allows running the same mapping algorithm across multiple jobs and across multiple engines. Running the same mapping algorithm across multiple engines requires a compatible external database. New APIs now support migrating mappings from existing mapping algorithms to the new mapping algorithms.

- **Algorithm replacement APIs**

APIs are now being introduced to list and replace algorithms.

Group	Endpoints	Description
algorithm	GET /algorithms/	Retrieves all usage of the algorithm specified in the request path.
algorithm	PUT /algorithms/	Updates all usage of the algorithm specified in the request path to use the new algorithm name supplied as a query parameter.

For more information, see [Managing Algorithm Usage](#).

Group	Endpoints	Description
algorithm	GET /algorithms/migration	Returns a list of result objects describing each possible migration. One object is returned for every algorithm on the engine that can be migrated.
algorithm	POST /algorithms/	Creates a new algorithm named newAlgorithmName (from the API query parameters), by migrating from the algorithm named in the query path.

For more information, see [Migrating algorithms](#).

- **New phone masking algorithm**

A new masking algorithm for the phone number framework for US and international numbers is now available. Migration from the old phone masking algorithm to the new one is required. For more information on transition, see [Delphix Community Post](#).

- **New custom SQL API**

In this release, Delphix has extended the list of API-endpoints by adding a new table-metadata endpoint for generating custom SQL for the given tableMetadataId.

The API endpoint is :

Group	Endpoints	Description
tableMetadata	GET /table-metadata/	Generates a custom SQL.

Release 6.0.9.0

- **Masking SDK driver support plugins**

The Masking SDK functionality is extended with the ability to develop a new kind of plugin, called driver support plugins. These allow the execution of developer-defined tasks as part of a masking job.

- **Masking SFTP connector is extended with a new flag UserDirIsRoot**

Delphix introduces a new flag, setting whether the SFTP Connector configured Path is relative or absolute.

- **New Email framework**

Delphix introduces a new Email Framework along with two default algorithm instances. This functionality allows for more customization in masking email addresses.

- **New copy environment API**

In this release, Delphix has extended the list of API-endpoints by adding a new API for copying environments.

The API endpoint is :

Group	Endpoints	Description
environment	POST /environments/	Copy environment objects in the same or a different application

Release 6.0.8.0

- **New name and full name frameworks**

Delphix introduces new Name and Full Name Frameworks, as well as their default algorithms instances. That functionality adds flexibility and more sophisticated ways for name masking.

- **Masking SDK multiple plugins capacity**

Masking SDK functionality is extended with an option of loading multiple plugins and chaining extensible algorithms based on different plugins. The dlpx-core plugin is uploaded by default.

- **New regex decompose algorithm chaining framework**

Delphix introduces the Regex Decompose extensible algorithm framework, which allows the capability to build new algorithms from a combination of predefined actions and existing algorithms.

- **Enclosure escape character support for delimited file masking**

In this release, Delphix has added escape character support for delimited file masking. Specifically the following were added:

- **Enclosure escaping strategy:** The user can configure the enclosure escape character from the UI/API to escape the enclosure. To configure the enclosure escape character from the UI, the user must select the "Enclosure Escaping Strategy" dropdown value as per the below options on the edit Rule Set popup window.
 - i. **Double Enclosure:** Double enclosure option will set the escape character value same as enclosure value.
 - ii. **Custom:** By selecting custom option, the user can specify any single character as an enclosure escape character, except the "escape sequences" and "control characters".

- **Escape "Enclosure Escape Character"**

The user can escape the "enclosure escape character" itself by clicking on the Escape "Enclosure Escape Character" checkbox on the edit RuleSet popup window.

For more detailed information, see [Managing rule sets](#).

Release 6.0.7.0

- **New date masking frameworks**

Delphix introduces new date masking frameworks, which includes date replacement, date shift, and multi-column dates. These new frameworks obviate the need for many of the custom date algorithms that were required in the past. Delphix also introduces new default implementations of common date-masking functionality. The new date masking frameworks are briefly described below.

- Date Replacement: Selects a replacement value from a configurable date range.
- Date Shift: Produces a replacement value by randomly shifting the input date by a configurable increment range.
- Multi-column Date: Masks date values that have a dependency, such as admission and discharge date using the same algorithm as Date Shift. This allows masking of both the initial date and the difference between the dates.

- **New credit card masking algorithms**

Delphix introduces a robust payment-card masking framework, as well as a default algorithm implementation for credit card data. The legacy credit card algorithm (that produced random values) is being replaced by the new default instance, which provides consistent masking results, a unique output for every valid input, always changes a valid input value, and preserves all non-digit portions of the input value.

- **Masking Engine changes for users and groups**

This enhancement adds stronger on-Masking Engine safeguards to the Users and Groups experience delivered in Central Management, in which the access to a Masking Engine's objects is determined by assigning authorization via global access groups. Specifically, when an engine opts into the global model, it relinquishes local control of object access. With this, the local enforcement of global (Central Management) settings is strengthened by deactivating local object access in the UI, thus ensuring the local values will not be overridden via frequent, periodic scans from Central Management.

- **New forgot and reset password APIs**

In this release, Delphix has extended the list of API-endpoints by adding two new API's related to the existing Forgot and Reset password feature for a user, which was available via GUI only till now.

The two new sets of API endpoints are :

Group	Endpoints	Description
user	POST /users/forgot-password	Send reset password mail to the user
	POST /users/reset-password	Reset new password for the user

The forgot-password API will generate and send a password reset link to the registered email id of the user, for which the password has to be reset.

The reset-password API will use the token sent via the password reset link, to set the new password.

- **Control character support for delimited file masking**

In this release, Delphix has added the control character support for delimited file masking. Specifically the following were added:

- a. **Control character as a delimiter:** The user can specify a control character as the delimiter from UI/API.
- b. **Control character as an end of record:** The user can specify a control character as the end of record from UI/API.
- c. **Control character as a value:** Delimited files containing values with control characters are now supported.

- **Date-Time format change for the API response**

In this release, the date-time format for API responses is changed

From: yyyy-MM-dd'T'HH:mm:ss.SSSZ e.g. 2021-03-17T17:35:39.352+0000

To: yyyy-MM-dd'T'HH:mm:ss.SSSXXX e.g. 2021-03-17T17: 35:39.352+00:00.

The API endpoints below will be affected by this change:

- GET /system-information
- GET /plugin
- GET /profile-jobs
- GET /profile-sets
- GET /execution-events
- GET /async-tasks
- GET /audit-logs
- GET /algorithms in algorithm extension object
- GET /execution-components
- GET /jdbc-drivers
- GET /masking-jobs
- GET /reidentification-jobs
- GET /tokenization-jobs

Release 6.0.6.0

- **Certifications**

This release adds support for DB2 iSeries v7.4.

- **Multi-column algorithm**

In this release, Delphix has introduced a Multi-Column Extensible Algorithm mechanism, which allows masking multiple columns of the same table conditional to their values (or using any other logic needed by the customer). To use the Multi-Column Algorithm Framework, users first create an algorithm via the Masking SDK and then install their algorithm on a Masking Engine via the Extensible Algorithm Plugin interface.

- **Latest API version**

The latest masking API version supported on the engine will be included in the `GET /system-information` API response.

- **Custom database connection properties**

There is now a way to specify custom connection properties for all of our database connector types by uploading a properties file. For more information, see [Database Connection Properties](#).

Release 6.0.5.0

- **Certifications**

This release adds support for the following certificates:

- MySQL 8
- Postgres SQL 12
- DB2 LUW 11.5
- Oracle Database Cloud Services on Virtual Machines
- Oracle Database Cloud Services on Bare Metal
- Google Cloud SQL for PostgreSQL
- Google Cloud SQL for MySQL
- Google Cloud SQL for SQL Server

- **Character mapping algorithm**

Delphix is introducing a replacement for the Segment Mapping Algorithm, the Character Mapping Algorithm. The new Character Mapping Algorithm is built using the recently released algorithm SDK, and in most common configurations this new algorithm will be faster and require less memory than the existing segment mapping algorithm. In addition, this new version does not have a length limitation for the input string and can handle non-ASCII characters.

- **Default API version**

Introducing the ability to specify the Masking API version to be used when the version is omitted from the base path of the Masking API request's URL.

- **New API version**

To reflect the API improvements mentioned above, the API version increased to 5.1.5 in this release. For a complete listing of version 5.1.5, see [Masking API Client](#).

Release 6.0.4.0

- **Certifications**

This release adds support for SQL Server 2017 and 2019.

- **Masking job memory improvements**

Memory management has been dramatically improved. Not only can jobs run with less memory, but the Masking Engine will also now ensure that jobs can only run if enough memory is available and that the engine cannot run out of memory.

Along with these changes, there are two new execution statuses: `CANCELLED` and `QUEUED`.

- **Extensible connector permissions change**

The first iteration of the Masking Extensible Connectors, supporting the ability to upload and use JDBC drivers, required that the permissions for each driver be enumerated at install time. Delphix has now replaced this mechanism with a fixed security policy blocking only the most dangerous permissions (specifically those that could inflict harm to the Masking Engine), removing the need for user management of permissions. It remains the case that the engine administrator must ensure that only trusted JDBC driver software is installed.

- **File masking performance**

The performance of file masking has been significantly improved.

- **Builtin extensible secure lookup framework**

Delphix has added a builtin, configurable Secure Lookup Algorithm Framework, based on the Extensible Algorithms feature (introduced in 6.0.3.0 release).

This framework provides better performance and new features when compared with the Legacy Secure Lookup Algorithms.

It allows configuring the case sensitivity of input values (true/false), and the case configuration of the output values:

```
Preserve Lookup File Case // i.e. as found in Lookup File Preserve Input
Case // i.e. preserve case of input value - UpperCase / LowerCase /
Mixed Force all Lowercase // forces output to LowerCase Force all
Uppercase // forces output to UpperCase
```

The algorithm instance (based on the new Secure Lookup Algorithm Framework) might be managed via the existing Algorithm API, similar to any other plugin algorithm. The GUI has been changed for configuring/editing Secure Lookup Algorithm. For more information, see [Secure Lookup Algorithm Framework](#).

- **Job scheduler removed**

As of this release, we have removed the Job Scheduler feature. The introduction of Masking's REST API several releases ago allowed customers to schedule job executions using their preferred job scheduler. As a result, the integrated scheduler is seldom used.

- **Free text redaction algorithm**

The redaction strategies used in a free text redaction algorithm have been renamed to "Allowlist" and "Denylist".

- **New API version**

To reflect the API improvements mentioned above, the API version increased to 5.1.4 in this release. For a complete listing of version 5.1.4, see [Masking API Client](#).

Release 6.0.3.0

- **Extensible algorithms**

We introduced a new, radically simpler, method to create new masking algorithms. With the new framework, Delphix partners and customers can create and share new algorithms.

Extensible algorithms and their related algorithm plugins can be managed through the following APIs:

Group	Endpoints	Description
plugin	GET /plugin	Get all plugins
	POST /plugin	Install plugin
	DELETE /plugin/	Delete plugin
	GET /plugin/	Get plugin detail by pluginId
	PUT /plugin/	Update plugin

Existing *algorithm* API is extended with the following endpoints:

Group	Endpoints	Description
algorithm	GET /algorithm/frameworks	Get all algorithm frameworks
	GET /algorithm/frameworks/id/	Get algorithm framework by frameworkId

- **UI-based environment sync**

Over the past several releases Delphix has introduced and refined the ability to synchronize objects between Masking Engines via the API. In 6.0.3, Delphix now supports importing and exporting environments via the UI.

Note: In this release, the deprecated XML import/export functionality has been removed. If you used the XML import/export feature in previous releases, you'll find the new Sync Environment feature to be a more robust and complete solution with complete API support in addition to being available in the UI.

- **New SQL Server JDBC driver**

The product switched from the jTDS JDBC driver to Microsoft's official open-source JDBC driver. This was done to obtain improved support for recent versions of SQL Server.

All SQL Server basic connectors will be converted transparently. If you used a SQL Server Advanced connector or a Generic connector using the jTDS driver, you will need to manually convert your JDBC URL to the Microsoft JDBC driver's format. To perform this conversion, see the references for the [jTDS parameters](#) and the [Microsoft JDBC parameters](#). Delphix Customer Support's upgrade validation checks will detect any SQL Server Advanced connectors and Generic connectors using the jTDS driver in your installation and they will notify you of the need to manually convert those connectors.

- **AzureSQL managed databases**

This release is certified to be compatible with the following Azure SQL Managed Databases:

- Azure Database for PostgreSQL service
- Azure Database for MySQL service
- Azure Database for MariaDB service
- Azure Database for SQL

Note: You must enable support for non-TLS connections.

- **File masking performance**

This release contains significant performance improvements for delimited and XML file masking.

- **New API version**

To reflect the API improvements mentioned above, the API version increased to 5.1.3 in this release. For a complete listing of version 5.1.3, see [Masking API Client](#).

Release 6.0.2.0

- **Certifications**

This release adds support for Oracle 19c.

- **Mainframe data set improvements for masking**

This release delivers multiple quality-of-experience enhancements around mainframe masking workflows:

- **Mainframe masking performance:** Anyone masking mainframe data sets may see a large improvement in performance.
- **Engine sync support for mainframe:** The Sync APIs and workflows now support mainframe objects: connectors, rule sets, jobs, and formats.
- **Mainframe data set record type APIs:** This enhancement builds upon the recent release of Record Type APIs to include mainframe support. You will now be able to manage Mainframe data set record types via REST API, including redefine conditions. When masking a mainframe data set, the Masking Engine uses a mainframe data set format to interpret the data set's contents. A mainframe data set format has one default record type "All Record". If a mainframe data set format contains redefined

fields, each redefined and redefines field will have a corresponding record type that holds the redefined condition for the redefined and redefines fields. Specifically, the following APIs were added:

Group	Endpoints	Description
mainframeDatasetRecordType	GET /mainframe-dataset-record-types	Get all Mainframe Dataset record type
	GET /mainframe-dataset-record-types/	Get Mainframe Dataset record type by ID
	PUT /mainframe-dataset-record-types/	Update Mainframe Dataset record type by ID

- For more information on redefine conditions, see [Managing a mainframe inventory](#).
- **JDBC to delimited files support**
On-the-fly masking jobs with a JDBC source and delimited file target are now supported. This is targeted at users with data lake applications. This is targeted at users with data lake applications who wish to extract unmasked data using a JDBC connection and insert masked data back using a bulk file load mechanism.
- **Environment sync support for masking**
With this release, an entire environment is now syncable with a single operation via the Sync REST APIs. Previously, Sync users would have to export/import objects on an individual basis, the process now is far more streamlined. Note: Environment Sync APIs are the preferred way of handling environment export/import versus XML-based transfer.
- **New API version**
To reflect the API improvements mentioned above, the API version increased to 5.1.2 in this release. For a complete listing of version 5.1.2, see [Masking API Client](#).

Release 6.0.1.0

- **Extended connectors**
Extended connectors is a new feature that allows you to upload additional JDBC Drivers to the Continuous Compliance engine. This enables masking data sources that are not natively supported by Continuous Compliance. For more information, see [Managing Extended Connectors](#).
- **Sync for tokenization and reidentification jobs**
The Sync feature allows you to coordinate the operation of multiple engines. This release adds Sync support for Tokenization and Reidentification Jobs. For more information on the Sync feature, see [Managing Multiple Engines for Masking](#).
- **File record type APIs**
When masking a delimited or fixed length file, the Masking Engine uses a file format to interpret the file's contents. Each format has one or more record types. In previous releases, these record types could only be created and managed through the graphical user interface. This release adds the ability to also create and manage file record types through the APIs. Specifically, the following APIs were added:

Group	Endpoints	Description
recordType	GET /record-types	Get all record type
	POST /record-types	Create record type

Group	Endpoints	Description
	DELETE /record-types/	Delete record type by ID
	GET /record-types/	Get record type by ID
	PUT /record-types/	Update record type
recordTypeQualifier	GET /record-type-qualifiers	Get all record type qualifiers
	POST /record-type-qualifiers	Create record type qualifier
	DELETE /record-type-qualifiers/	Delete record type qualifier by ID
	GET /record-type-qualifiers/	Get record type qualifier by ID
	PUT /record-type-qualifiers/	Update record type qualifier by ID

Note that record types are only used for delimited and fixed-length file formats. For more information on record types, see [Adding Record Types for Files](#).

Release 6.0.0.0

- **Objects names requirements**

Delphix 6.0 adds validations for object names that can be created/renamed manually. For more information, see [Naming Requirements](#).

Please note that enforcing these requirements might fail the import, sync, or upgrade from pre-6.0 release.

For resolving those failures, see [Knowledge Base Article KBA5096](https://support.delphix.com/Delphix_Masking_Engine/Object_Naming_Requirements_(KBA5096)).

- **Versioning framework**

6.0 marks the release of version 5.1 of the Masking API. For information on how the Masking API is versioned, see [Masking API Versioning Documentation](#).

- **New API endpoints**

In 6.0 we have expanded the list of API endpoints to include:

Group	Endpoints	Description
Application	DELETE /applications/	Delete application by ID
Mount Filesystem	GET /mount-filesystem	Get all mounts
	POST /mount-filesystem	Create a mount
	GET /mount-filesystem/	Get a mount by ID

Group	Endpoints	Description
	DELETE /mount-filesystem/	Delete a mount by ID
	PUT /mount-filesystem/	Update a mount by ID
	PUT /mount-filesystem/	Connect a mount by ID
	PUT /mount-filesystem/	Disconnect a mount by ID
	PUT /mount-filesystem/	Remount a mount by ID

In addition to the new API endpoints, we have improved existing API endpoints. These improvements include:

- Addition of the applicationId field to the application model
 - Replacement of the application field with an applicationId field in the Environment model
 - Removal of the classification field from the domain model
 - Addition of the rulesetType field to the Masking, Profiling, Reidentification, and Tokenization job models.
 - Addition of mountName in the ConnectionInfo of a file connector and a mainframe dataset connector to use a filesystem mount point.
 - For more information on Continuous Compliance APIs, see [API documentation](#).
- **NFS and CIFS mounts**
In previous releases, the Masking Engine has supported masking files via FTP or SFTP. In this release, we have added the ability for users to directly mount and mask a file system over NFS and CIFS. This should dramatically simplify the process of file masking. As with other Masking Engine objects, the Sync feature can be used to coordinate mount objects across multiple engines. For more information on the mount feature, see [Managing Remote Mounts](#).

Release 5.3

- **Synchronizing masking jobs and universal settings across Engines**
In 5.2 we introduced the ability to synchronize Masking Algorithms between engines to ensure consistent masking, regardless of the engine executing the masking. In 5.3 we are expanding the list of syncable objects to include:
 - Masking Jobs
 - Connectors
 - Rulesets
 - Domains
 - File Formats

The sync of objects is possible through improvements to several sync API endpoints, including:

- GET /syncable-objects[?object_type=
- POST /export
- POST /export-async
- POST /import
- POST /import-async

This expansion of syncable objects ensures that users can sync their Masking Jobs and all the objects necessary for that masking job to execute successfully - regardless of the masking engine it lives on, allowing for easier scaling of Continuous Compliance across the enterprise. For more information, see [Managing Multiple Masking Engines](#).

• **Support for Kerberized connections**

In 5.2.4 we added support for Kerberos for our Oracle Masking Connector. In 5.3 we have expanded the list of connectors that support Kerberos to:

- SQL Server
- Sybase
- To enable Kerberized connectors your engine must be configured properly and you must configure your masking Connectors for Kerberos. Kerberos can be enabled by going to the Advanced mode on Oracle, SQL Server and Sybase. For more information, see [Managing Connectors](#).

• **New API endpoints**

In 5.2 we released an all-new set of API endpoints allowing for the automation of many masking workflows. In 5.3, we have expanded this list of API endpoints around Algorithms, Users, Roles, File Upload, System Information, Login, Rulesets, and Connector. Below are the net new API endpoints:

Group	Endpoints	Description
Algorithms	POST /algorithms	Create algorithm
	DELETE /algorithms/	Delete algorithm by name
	GET /algorithms/	Get algorithm by name
	PUT /algorithms/	Update algorithm by name
	PUT /algorithms/	Randomize key by name
Users	GET /users	Get all users
	POST /users	Create user
	DELETE /users/	Delete user by ID

Group	Endpoints	Description
	GET /users/	Get user by ID
	PUT /users/	Update user by ID
Roles	GET /roles	Get all roles
	POST /roles	Create role
	DELETE /roles/	Delete role by ID
	GET /roles/	Get role by ID
	PUT /roles/	Update role by ID
Rulesets	PUT /database-rulesets/	Update the rule set's tables
	PUT /database-rulesets/	Refresh the rule set
Connectors	POST /database-connectors/	Test a database connector
	POST /database-connectors/test	Test an unsaved database connector
	POST /file-connectors/	Test a file connector
	POST /file-connectors/test	Test an unsaved file connector
Async Tasks	GET /async-tasks	Get all asyncTasks
	GET /async-tasks/	Get asyncTask by ID
	PUT /async-tasks/	Cancel asyncTask by ID
File Upload/Download	DELETE /file-uploads	Delete all file uploads
	POST /file-uploads	Upload file
	GET /file-downloads/	Download file

Group	Endpoints	Description
System Information	GET /system-information	Get version, etc.
Login/Logout	PUT /logout	User logout
Executions	GET /execution-components	Status for a table, file, or Mainframe data set
Tokenization Job	GET /tokenization-jobs	Get all tokenization jobs
	POST /tokenization-jobs	Create tokenization job
	DELETE /tokenization-jobs/	Delete tokenization job by ID
	GET /tokenization-jobs/	Get tokenization job by ID
	PUT /tokenization-jobs/	Update tokenization job by ID
Re-identification Job	GET /reidentification-jobs	Get all re-identification jobs
	POST /reidentification-jobs	Create re-identification job
	DELETE /reidentification-jobs/	Delete re-identification job by ID
	GET /reidentification-jobs/	Get re-identification job by ID
	PUT /reidentification-jobs/	Update re-identification job by ID
Database Rulesets	PUT	Update Database Ruleset by ID

In addition to the net new API endpoints, we have improved pre-existing API endpoints. Some of the improvements include:

- Addition of DB2 iSeries and Mainframe to connector endpoints.
- Addition of Kerberos configuration on Oracle, SQL Server, and Sybase connectors
- Ability to have ruleset refresh drop tables
- Support for XML file types
- Addition of dataType to column metadata
- Addition of isProfilerWritable field to file-field-metadata endpoints. This is now represented in the API as a new *isProfilerWritable* boolean field in the body of a file-field-metadata. When the isProfilerWritable field is set to true, the algorithm/domain assignment on a column can be overwritten by the profiler. When the field is false, it may not be overwritten.
- Addition of multipleProfilerCheck field to Profile Job endpoints. This feature is turned on using the boolean field in the body of a profile job. The job profiler normally stops profiling a column as soon as

it flags a field as sensitive. If *multipleProfilerCheck* is true, the profiler will continue to scan the column for additional sensitive patterns. In the event that it finds more than one pattern, it will tag all the data domains found and apply 'one' standard algorithm for all those domains. The standard algorithm is 'Null SL' as of 5.3.4.0. This feature was formerly called 'multi PHI'.

For more information on Continuous Compliance APIs, see [API documentation](#). Please note that the previous generation of Masking APIs (commonly referred to as V4) is EOL and no longer supported in this release. All users are encouraged to migrate to the V5 APIs.

Fixed issues

Release 13.0.0.0

Bug Number	Description
DLPX-79868	Fixed an issue with Inventory export/import functionality to avoid CSV injection.
DLPX-82217	Fixed an issue where <code>dlpx-core:FirstName</code> would not mask as expected.
DLPX-82912	Fixed an issue showing HTML code instead of an actual success/error GIF when performing an inventory import from the UI.
DLPX-84783	Fixed an issue with permissions ambiguity when viewing Jobs logs vs Application logs. Now, a user who can view a job on the Monitor page will be able to view logs as well. In addition, a non-admin user with permission to view diagnostics will be able to view application logs.
DLPX-84875	Fixed an issue where a user with create/edit permissions was not able to create or edit the classifier.
DLPX-85564	Fixed an issue where non-admin user access was restricted to only the tasks they started (API methods: <code>GET /async-tasks</code> and <code>GET /async-tasks/{asyncTaskId}</code>).
DLPX-86134	Fixed an issue where a masking algorithm migration failure during engine upgrade causes the Algorithm UI to not load and the <code>dlpx-core</code> plugin to not upgrade.
DLPX-86254	Added Audit log entries for rulesets and jobs in case of the deletion of a related connector.
DLPX-86715	Secured GraphQL APIs by adding Content-Security-Policy headers and removing unsafe CPRS headers.
DLPX-86717	Fixed an issue where ASDD jobs on Maria DB will no longer fail with SQL exceptions due to incorrect identifier quoting.

Security Fixes

Bug Number	Introduced	Description	Security Bulletin
------------	------------	-------------	-------------------

DLPX-86715	13.0.0.0	Secured GraphQL APIs by adding Content-Security-Policy headers and removing unsafe CORS headers.	TB110
------------	----------	--	-------

Release 12.0.0.0

Bug Number	Description
DLPX-85600	Fixed the DROP INDEX feature for MySQL databases.
DLPX-85864	Fixed an issue related to uninformative masking progress whenever the row count is more than 2.1bn for SQL server.
DLPX-85971	Fixed an issue where a masking job fails if column name contains special characters and custom SQL is used.
DLPX-86049	Fixed an issue related to the GraphQL service returning Http-502 in certain scenarios.
DLPX-86196	Fixed an issue where segment mapping algorithms exported using Engine Synchronization (from releases 6.0.15.0 -> 10.0) were not validated upon import to 11.0.
DLPX-86249	Updated protobuf-java dependency version to 3.23.2.
DLPX-86259	ASDD job execution will now correctly show execution event(s) in the UI when failures occur.

Security fixes

Bug Number	Introduced	Description	Security Bulletin
DLPX-86329	6.0.13.0	Sysadmin can execute shell commands on the underlying Operating System.	TB109

Release 11.0.0.0

Bug Number	Description
DLPX-85961	Fixed an issue where the Filter, Logical Key, and Custom sql for DB tables could not be updated when using the table name search feature inside DB rulesets.
DLPX-85867	The non-conformant data will no longer get copied to all the tables.
DLPX-85862	Fixed a CSS issue causing the Inventory UI to slide to the bottom of the page.
DLPX-85743	Fixed missing tables in the UI-Job Monitor-Waiting grid page.
DLPX-85514	Fixed an issue in the ASDD profiler where the count of rows profiled was not displayed in the job monitor.
DLPX-85459	Fixed an issue that sometimes caused long delays before log messages related to jobs would show in the main logs.
DLPX-86112	ASDD profiling fails when multiple tables with similar names containing an underscore are present.
DLPX-82007	The deletion of a user-created mapping algorithm now deletes associated mappings from MDS.
DLPX-85289	Added changes for ensuring JDBC parameters are handled properly with DB2 mainframe connectors.

Security fixes

Bug Number	Introduced	Description	Security Bulletin
DLPX-85414	6.0.15.0	Unmasked Data or Masking Job Failure When Using a Migrated Version of the Segment Mapping Algorithm	TB108

Release 10.0.0.0

Bug Number	Description
DLPX-79335	Fix for Oracle databases masking job failing while masking XMLType columns, having masked data more than 2000 characters.
DLPX-84025	If a mapping algorithm is used on a field that contains only characters set to be ignored, a non-conforming data error will no longer be thrown.
DLPX-84303	Phone Unique and Phone US algorithms no longer report non-conforming data as an error, since masking can continue and it can obscure true breaking errors.
DLPX-84874	Fixed an issue that causes long load times for the algorithm and domain settings screens.
DLPX-84966	Prevents masking job failures by adding extra validation when defining alphanumeric segments with Segment Mapping algorithms.
DLPX-85382	Fixed an issue with the ASDD profiler that can make data-level profiling take much longer than expected on large tables in a Microsoft SQL Server DB.
DLPX-85437	Fixed an issue that can make deletion of a database ruleset very time consuming when many profiler results exist.
DLPX-85961	

Release 9.0.0.0

Bug Number	Description
DLPX-82535	Added an application setting to change the default API page size for all GET APIs from 1 to 5000.
DLPX-84110	Fixed an issue where the drop index checkbox reset to an unchecked state after the masking dialog box is closed.
DLPX-84336	The search options on the Environment and Ruleset page now allow for "-" character matching.
DLPX-84488	Phone Unique and Phone US algorithms saw a failure on zero width and formatting characters included with a phone number in a database row. Phone algorithms can now pass over these non-numeric characters as expected.

Release 8.0.0.0

Bug Number	Description
DLPX-73334	Profiling jobs in tokenization environments where domains lack a tokenization algorithm assignment no longer fail.
DLPX-82175	Fixed an issue where editing a DB ruleset was taking a long time to modify the logical key of a table with a massive amount of data, by moving the validation from the UI to a backend check for null values in a conditional column.
DLPX-83340	Encryption algorithm upgrade for Engine sync bundles.
DLPX-83536	The city column level profiling expression has been updated to not match "address".
DLPX-84046	Added a flag 'trimWhitespaceFromInput' to the Secure Lookup algorithm that can be used to restore pre-6.0.12.0 trim behavior.
DLPX-84081	Fixed an issue where the Continuous Compliance Engine environment revisionHash changes on profile job execution, even though there was no change in inventory.
DLPX-84230	Fixed an issue where support for moving data from DB tables to delimited files was inoperable.
DLPX-84463	Swagger-UI minor version upgrade for bug fixes done in library.
DLPX-84525	Inventory page fails to load on Continuous Compliance Engine when there are more than 32,767 masked columns on the engine.

Release 7.0.0.0

Bug Number	Description
DLPX-76693	Added an execution-event and a log error message showing the missing column names, when custom SQL is used in the ruleset and some columns are not specified in the query.
DLPX-79530	Continuous Compliance Engine environment revisionHash changes frequently due to execution of masking and tokenization jobs.
DLPX-80124	Column name is checked if it contains a space, if a true rename of the column name is set to false, so that the column name is not renamed. Note, this change is only for column names having whitespace and not for columns having special characters.

Bug Number	Description
DLPX-81503	Fixed an issue when wrong environment link was getting generated on Monitor page.
DLPX-82950	Fixed an issue where Filter By in the Inventory page on the UI was intermittently working for File inventories.
DLPX-83035	In case of Job failures, Error details will be displayed in Status/Logs dialog on the Execution details page.
DLPX-83566	Fixed an issue when connection to NFS mount failed after the NFS server was restarted.
DLPX-83593	Fixed an issue where unnecessary locking can cause slowdowns and possibly deadlocks.
DLPX-83659	An update was made to the DB2 license upload script to make it compatible with all Continuous Compliance versions 6.0.14.0 and later.
DLPX-83804	Fixed an issue where assigning Dataset File Formats using the API was not working for some formats.
DLPX-83809	Fixed an issue causing random job failures when masking SQL Server tables that have columns of date/time data type as part of the PK.
DLPX-83817	Fixed an issue where non-admin users can submit a inventory change for approval workflow without approval inventory permission.
DLPX-83930	Fixed an issue that caused some environment sync import operations to fail with PersistentObjectException.

Release 6.0.17.0

Bug Number	Description
DLPX-46230	Fixed an issue where tables with a Primary Key column of datatype RAW would cause an error when selected for masking.
DLPX-55224	Search using wildcards and substrings is now allowed on search boxes across multiple pages.
DLPX-69096	Email addresses are case sensitive when SSO is enabled.

Bug Number	Description
DLPX-69501	User defined domains cannot be deleted when assigned to profiler expressions.
DLPX-76315	Display of field level redefine fields have been applied in the Inventory screen.
DLPX-76696	Audit logs being displayed is not limited to 1000 entries, all the logs recorded will be displayed with a pagination of 50 rows per page. API of audit logs is modified to include all search and filter parameters used in UI.
DLPX-77069	Fixed an issue where using regex for File Name Patterns in xmlfile Rule Sets would succeed but give an error.
DLPX-80496	Removed extraneous link to the job's execution log from the monitor screen for profiling jobs.
DLPX-80539	Fixed an issue where some failures detected during file masking job generation are not correctly reported via an execution event.
DLPX-80540	Fixed an issue where some failures during file masking job generation leave a job in running status indefinitely.
DLPX-80984	Fixed negative row counts when custom SQL includes column name that contains SQL reserved word.
DLPX-81311	A generic DB error message will now be shown to users for fetching and creating rule set.
DLPX-81725	Fixed a bug that could cause the masking engine to run out of memory processing results from large jobs.
DLPX-82064	Fixed MSSQL masking performance issue when using Kerberos authentication and masked columns that are unicode, but primary keys are non-unicode.
DLPX-82289	Fix provided to support backslashes and other special characters that might be required to form a valid regular expression.
DLPX-82310	User's first and last names are now redacted in the support bundle.
DLPX-82579	Improved performance to prevent GUI lag when navigating across pages.
DLPX-82864	Error message for invalid LDAP authentication attempt has been updated to prevent username harvesting.

Bug Number	Description
DLPX-82925	Fixed an issue where the Profile Results screen breaks due to a database error.
DLPX-83026	Fixed an issue causing the DateShiftDiscrete algorithm to not be assigned a new random key when an engine is deployed.
DLPX-83086	Upgraded Swagger UI to enhance the security and usability of the swagger API Client.
DLPX-83200	Fixed scale issues related to storing job logs in the product's internal database.
DLPX-83232	Added an option to the Secure Lookup algorithm to disable whitespace cleanup when the lookup file is loaded.
DLPX-83354	Updated the profile expression 'Full_Name_V2' to reduce the number of false positive results. This change only applies to newly deployed Continuous Compliance Engines.
DLPX-83427	Fixed Directory Travel vulnerability for the export inventory UI.
DLPX-83431	Upgraded Apache Commons to 1.10.0.
DLPX-83576	Fixed an issue where the Support bundle process did not collect the /etc/hotfix file.

Release 6.0.16.0

Bug Number	Description
DLPX-49116	Data truncation when masking CHAR(n) using Segment Mapping to mask short numeric segments
DLPX-73539	Pdf for Audit log does not contain Status column
DLPX-77777	View only connector privilege user will be able to see connection details and test connection using Masking UI
DLPX-77929	Algorithm API: Add mask_type attribute in the Api response for component type algorithm.
DLPX-78474	Able to validate algorithm from: Algorithms > edit extended instance > "validate configuration" button

Bug Number	Description
DLPX-79358	Monitor page UI Start and End Date Filters do not work as expected
DLPX-79607	Show a better user friendly error message when creating full name algorithm with invalid input
DLPX-79743	User will not be logged out, after performing an operation for which they have insufficient privilege. API will be throwing 403 status code, instead of 401.
DLPX-80304	isidentity has been added to the GET /column-metadata API response that indicates whether the table column is an identity column
DLPX-80668	For Job execution steps, icon for last event step "Job completed" will be aligned to final status of the job. (1) If job is success - Green tick icon (2) If job has failure - Red failed icon, (3) If job was cancelled - Red cancelled icon
DLPX-80784	Addition of new application setting group to drive strict content security policy
DLPX-81058	Saving File/Copybook field properties on Inventory screen will now retain the ruleset selected
DLPX-81422	Updated Job Monitor page with small UI fixes. Page position will remain same after page refresh. Renamed job execution events.
DLPX-81730	Segment Mapping pads short numeric input even when not masked
DLPX-81807	Fixed an issue that can very rarely cause the masking service to fail with a stack overflow during startup
DLPX-81895	Data profiling results are not shown under job -> monitor page -> results tab for delimited, fixed and XML files
DLPX-81897	Internal error when GET syncable-objects API is called with object_type=LOOKUP
DLPX-81935	Profiling regex with escaped double-quotes fails in compiling javascript
DLPX-81946	Allow masking of dates containing month values in all-upper character case
DLPX-82025	Setting pseudo column as ROWID in the Logical Key of the Ruleset is possible
DLPX-82057	GET /file-uploads API endpoint returns a 500 error when plugins exist on the engine

Bug Number	Description
DLPX-82086	From this change, all the timestamps on Masking UI will be as per user timezone. Logs, search filter on Job monitor and Audit page will be in UTC Timezone. User won't be able to change timezone on UI. It will be by default as per their browser timezone.
DLPX-82166	Segment Mapping v2 UI cannot specify SPACE as an ignore character
DLPX-82186	Clicking outside connector dialog will not dismiss the dialog
DLPX-82337	Segment Mapping algorithm with CONSTANT segment should not be allowed in Tokenization job
DLPX-82355	Fixes an internal server error that occurred when attempting to import a legacy Secure Lookup algorithm of type LOOKUP.
DLPX-82491	From this change, User can hover on the time displayed on UI and will able to identify timezone offset related information.
DLPX-82627	Fix an incorrect check preventing the Segment Mapping Algorithm from running in REIDENTIFY mode

Release 6.0.15.0

Bug Number	Description
DLPX-77088	Fixed an issue causing the Full Name Algorithm to fail masking when a single non-alphanumeric char is present, one of the input words is in the particles file, and the 'Last-First-Middle' convention is used.
DLPX-79547	Fixed an issue causing: <code>MSSQL Masking Job XmlArtistFailureException: Input has a cycle and cannot be used with XML Artist.</code>
DLPX-80123	Non-conforming Data is now reported when a mapping algorithm's available mappings are exhausted.
DLPX-80225	A new interface has been developed in the masking SDK for SingleOperationTask.
DLPX-80603	Fixed an issue causing read-only multi-column fields to be upgraded incorrectly.
DLPX-80604	Fixed an issue where records are truncated if the field length is not '0' in the fixed width file format, and 'whole file masking' selected.

Bug Number	Description
DLPX-80732	Fixed incorrect text in an error message that occurs while exporting a profile_typed_expression.
DLPX-80788	Added a check to validate same site cookies transfer.
DLPX-80837	Fixed an issue where a copybook with a file name more than 24 characters fails loading: "String field too long".
DLPX-80842	Fixed an issue where the "All Fields" button in database inventory unexpectedly refreshes inventory for the displayed table.
DLPX-80922	Changed the version for the extensibility API to 1.10.0.
DLPX-81067	Cleaned up some stale build properties.
DLPX-81110	Fixed an issue where the filter on the Job Wizard Inventory screen was missing.
DLPX-81128	Fixed an issue of being unable to create/edit a profile set from the GUI.
DLPX-81175	Fixed an issue with Tokenization v2 instances where character mapping fallback and the same character groups produced the same results within a job, and inconsistent results across jobs.
DLPX-81259	Removed the 256 character limit on masking copybook redefine conditions.
DLPX-81261	Fixed an issue where masking VSAM with different Algorithms causes: <code>IndexOutOfBoundsException: Index: 0, Size: 0.</code>
DLPX-81330	Permanent file upload is now available for files that are not explicitly associated with a JDBC driver, algorithm, driver support plugin, or connection properties.
DLPX-81397	Fixed an issue where the JOB ID was not displayed in Job Monitor Processing or Waiting tabs.
DLPX-81512	Fixed an issue where sync bundles from releases prior to 6.0.15.0 could not be imported into 6.0.15+ releases.
DLPX-81666	Fixed an issue where sync exports of environments/jobs/connectors using non-default driver support settings were unusable. This happened due to failures of jobs/edits on the system where they are imported, where the non-default driver support settings are used.

Release 6.0.14.0

Bug Number	Description
DLPX-56200	Stopped execution.job.jobId logs from spamming the log files.
DLPX-77999	The Legal disclaimer and description of the framework on the Algorithm GUI now appear.
DLPX-78560	The issue with sync compatibility for multiple headers and footers is now fixed.
DLPX-78671	Added pagination for Syncable objects with the object_type LOOKUP.
DLPX-78850	Fixed an issue where the mainframe file format delete fails with NoSuchFileException if the format file is not present on disk.
DLPX-79160	Fixed an issue that could cause sync import to fail due to inconsistent multi-column algorithm assignments in the sync document.
DLPX-79472	Fixed an issue that prevented the saving of multi-column algorithm masking assignments in file inventories.
DLPX-79608	Fixed an issue causing SLv2 to fail when the lookup file contains just spaces.
DLPX-79720	Fixed an issue where DriverSupport logs stopped printing after switching to the next log file due to file size limit.
DLPX-79865	Improvements have been made to the Forgotten Password API error message.
DLPX-79966	PostgreSQL driver updated from 42.2.23 to 42.3.2 version.
DLPX-79986	The expression_name field in the profile-type-expression endpoint has been renamed type_expression_name.
DLPX-80386	Column and data level profiler expressions are now tested in a predictable order - alphabetically, by expression name.
DLPX-80411	Upgraded Spring framework version from 5.2.5 to 5.2.20.
DLPX-80078	The issue while removing files with complex file permissions on EBS is now fixed.
DLPX-81071	HTML escape the inventory notes field.

Bug Number	Description
DLPX-81082	Addresses the issue described in TB098 .

Release 6.0.13.0

 This release renames Delphix Masking to Continuous Compliance.

Bug Number	Description
DLPX-77075	The issue with masking MSSQL date field that was causing the error "Conversion failed when converting date and/or time from character string" is now fixed.
DLPX-78363	This release adds support for API to get execution logs.
DLPX-78366	This release adds the new API endpoints to get execution component logs.
DLPX-78472	This release adds the total job time to the Masking job reports.
DLPX-78755	The issue with the failure of async export (if data is huge) with the OOM(Requested array size exceeds VM limit)error message is now fixed.
DLPX-78948	Previously, the edit user dialog was not opening for non-admin users if SSO is enabled. This issue is now resolved.
DLPX-79152	Extended algorithm enclosure handling was throwing an NPE when there are too few fields in a delimited file. This issue is now fixed.
DLPX-79177	The mapping Algorithm was failing in some cases(such as delimited files and BigInt Column) with error Mapping output value exceeding maximum value length of 0 characters. This issue is now fixed.
DLPX-79567	Previously drop index was failing if an index with the same name existed on the masked columns across multiple tables for MSSQL databases. This issue is now fixed.
DLPX-79627	The issue with the failure of the Masking SQL Server when special characters(such as] [or ') are used in the table column is now fixed.
DLPX-79632	The issue with the failure of the Masking Oracle DB when special characters are used in the table or column name is now fixed.

Bug Number	Description
DLPX-79803	The issue with the failure of the Masking with MSSQL database if table name contains '[' is now fixed.
DLPX-79804	The issue with the failure of the Masking with MSSQL database if table name contains '\' is now fixed.

Release 6.0.12.0

Log4j updates

Based on detailed testing and analysis, all the currently supported products are not susceptible to known log4j vulnerabilities. Please refer to [TB095 Technical Bulletin](#) for more information. All instances of log4j in currently supported Delphix products are updated to **log4j 2.17.1** as of this release.

Delphix keeps you updated on the latest developments and keeps releasing hotfixes, procedures, and workarounds for such critical vulnerabilities. For more information on how Delphix supports our product and customers in such cases, see [Delphix Product Security](#).

For more information, refer to the following pages:

- [TB095 log4j Vulnerabilities](#)
- [Uninstalling the Delphix Connector Service from the Target Database Servers](#)
- [Delphix Product Lifecycle Policies](#)
- [Product Security](#)

Fixed Issues

Bug Number	Description
DLPX-48506	The issue with the VSAM masking job failing with an error message, "Multiple entries with the same key: FILLER" is now fixed.
DLPX-64060	For the "Define Fields" popup in File Inventory, the previously saved algorithm is now displayed as selected. If domain and algorithm were not assigned, then selecting a domain will not select the respective default algorithm in the algorithm field.
DLPX-67419	The issue with the generation of the Generic Security Services API exception when performing data-level profiling on a Kerberized database is now fixed.
DLPX-69263	The issue with the failure of masking Hana DB using an extended connector when binary columns are masked or present for OTF jobs is now fixed.
DLPX-75726	The issue with the clearing of the file format configurations when modifying the file masking pattern is now fixed.

Bug Number	Description
DLPX-76752	Time format now includes seconds on the Monitor page for a better user experience.
DLPX-77036	The issue with setting Null for owner_id on referenced objects when deleting a user and resulting in NPEs is now fixed.
DLPX-77145	The issue with being unable to run any jobs - NPE in getTotalXmxOfRunningExecutions is now fixed.
DLPX-77166	Extended algorithms that support tokenization are now available to assign as the tokenization algorithm in domains.
DLPX-77233	PostgreSQL JDBC driver is upgraded to version 42.2.23.
DLPX-77258	This release fixes a bug in Data Level profiling when the specified schema is not the user's default schema.
DLPX-77401	The issue with not being able to extract the unmasked fields using API is now fixed.
DLPX-77502	This release now adds an end-point (POST) for file-field-metadata API.
DLPX-77503	Inventory GUI now uses a POST API end-point.
DLPX-77506	The issue with the failure of Data level profiling if the EnableDataLevelCount application is set to True is now fixed.
DLPX-77521	The masking engine now bars multiple headers and trailers for the record type.
DLPX-77524	This release adds filters to the table-metadata API.
DLPX-77594	The issue with a regular user not being able to submit an inventory change is now fixed.
DLPX-77629	This release changes the field labels from 'Prescript' and 'Postscript' to 'Pre SQL Script' and 'Post SQL Script' respectively in the Masking Job UI.
DLPX-77636	Job execution API now provides a job status filter to enhance the user experience.
DLPX-77688	The Character Mapping Algorithm's non-editable preserve range when editing the algorithm is now fixed.

Bug Number	Description
DLPX-77718	Users will now be able to associate a new parameter 'Whole File Masking' for any files listed on the Fixed File Rule Set page.
DLPX-77720	The issue with the displaying of an error message, "java.lang.NumberFormatException" when using Save & View option during the environment copy operation is now fixed.
DLPX-77767	Previously, the Delphix Masking engine used the incorrect HTML response code of 400 (Bad Request) for objects that could not be manipulated because they were currently in use. This release changes that to code 409 (Conflict).
DLPX-77786	This release blocks the creation of multiple header/trailer record types.
DLPX-77869	The issue where DESC order indexes were not being dropped and re-created as part of the Oracle Drop Indexes task has now been resolved. Functional indexes, including DESC order indexes, are now dropped and re-created on Oracle tables that contain any masked columns.
DLPX-77931	This release adds a translator to support Backward compatibility for PUT /file-field-metadata/
DLPX-77962	Users will now be unable to update fields like position and length for a fixed-width file if the 'Whole File Masking' feature is enabled.
DLPX-77963	For any Fixed-Width file, if the 'Whole file masking' option is selected, then Kettle Reads the complete content of the file and passes it as one single record to the configured algorithm.
DLPX-77976	This release replaces all the "NULL" values for the user_id column of the algorithm table by the ID of a Delphix internal user called 'deleted-user'.
DLPX-78105	Users will now see a proper error message when creating/updating the mainframe field if the provided date format is invalid.
DLPX-78116	This release adds an 'istokenizationSupported' flag in the Algorithm API response.
DLPX-78161	Created a function that uploads files bypassing tomcat's /tmp directory.
DLPX-78422	The issue with the logical key not being added to the table in Rule Set via GUI if the user is not the schema owner is now fixed.

Bug Number	Description
DLPX-78615	The issue with masking job throwing an exception while logging certain messages from plugin algorithms or driver support modules (This issue resulted in job deadlock during cleanup) is now fixed.
DLPX-78680	This release performs a clean-up of an obsolete lookup file attachment after making the import of an FTR-v2 algorithm.
DLPX-78740	The issue with the changing of an algorithm key when making the import of an FTR-v2 algorithm is now fixed. This release keeps the algorithm key unchanged.
DLPX-78743	This release updates all the masking dependencies on the Apache log4j library to version 2.15.0.
DLPX-78864	This release updates Log4j to version 2.0.17.
DLPX-78943	This release updates the log4j version to 2.17.1.

Release 6.0.11.0

Bug Number	Description
DLPX-55595	The issue where the Edit job dialog closes and leaves the screen greyed out with no errors while jobs are running has now been resolved.
DLPX-55595	The issue where the Edit job dialog closes and leaves the screen greyed out with no errors while jobs are running has now been resolved.
DLPX-65971	The issue where Cancel Masking job fails with "Execution status must be RUNNING, but is SUCCEDED" has now been resolved.
DLPX-67558	The issue where a Masking job appears to hang when masked columns are unicode, but the primary keys are non-unicode, has now been resolved.
DLPX-69778	SAML response should no longer be logged on successful SSO login.
DLPX-70104	Enhanced the date format validation for file-field-metadata and mainframe-dataset-field-metadata API.
DLPX-70499	The Monitor Page will now show an informative message if no jobs are returned.

Bug Number	Description
DLPX-72196	The issue when editing column properties for a file based inventory with no value selected for the ID Method field causing no validation to show has now been resolved.
DLPX-73326	There was an issue where when copying an environment, a dialog box shows a message that passwords will not be saved for connectors, but in the copied environment, the password information is present and Test Connection succeeds without any change. This issue has now been resolved.
DLPX-74245	The issue with inconsistent deletion behavior for a referenced database, file, and dataset connectors has now been resolved.
DLPX-74745	The <code>DEFAULT_MULTIPHI_ALGORITHM</code> application setting has been renamed to <code>DEFAULT_MULTIPLE_PROFILER_EXPRESSION_ALGORITHM</code> .
DLPX-75948	The issue showing inconsistent breadcrumbs for the VSAM/Mainframe Inventory screen has now been resolved.
DLPX-76365	The issue where the Trans Level Info table grows without bound has now been resolved.
DLPX-76574	The issue causing a failure to retrieve the ERROR or Warning column type has now been resolved.
DLPX-76678	Added validation to disallow null values in the logical key columns at the time of create or update.
DLPX-76707	The issue where update algorithm shows an error with, "installed by the plugin [plugin name], cannot be modified independently" has now been resolved.
DLPX-76847	The issue where Masking PK on Oracle adds ROWID to SELECT but uses PK in UPDATE has now been resolved.
DLPX-76931	The issue where the Masking UI strips extra characters from connector hostname when hostname exceeds max character limit has now been resolved.
DLPX-77056	The ruleset deletion validation message has been updated.
DLPX-77075	The issue where masking an MSSQL date field caused the error, "conversion failed when converting date and/or time from character string" has now been resolved.

Bug Number	Description
DLPX-77103	The issue where mixing extensible algorithms and mapplets in a VSAM jobs causes the job to crash has now been resolved.
DLPX-77138	The issue where the use of Carriage return \r breaks the inventory page when used in mainframe redefine condition has now been resolved.
DLPX-77139	The issue where <code>V2021_04_05_2__fix_algorithm_plugin_metadata</code> migration may fail with a "FileNotFound" exception has now been resolved.
DLPX-77159	The issue with VSAM Unmasked fields being truncated when redefines are present and an algorithm returns non-null results for null input has now been resolved.
DLPX-77267	The issue where an XML masking job can hang when GSSAPIAuthentication is enabled on the sftp server has now been resolved.
DLPX-77542	The issue where an extended connector SQL count fails when the column name contains the word 'FROM_DATA' in custom SQL has now been resolved.
DLPX-77544	The issue where deleting a masking user causes the deletion of the masking users' objects (meaning potential loss of important information, including historical information) has now been resolved.
DLPX-77710	The issue with a missing index on an Oracle DB after a successful masking job run has now been resolved.

Release 6.0.10.0

Bug Number	Description
DLPX-59886	You can now set a timeout for the FTP connections.
DLPX-70680	The issue with the increasing of the JobLogs without bounds has now been resolved.
DLPX-71259	Masking Oracle LONG RAW length is now set to 0 characters.
DLPX-71993	The need for the 'Repository' on the Masking Monitor page is now removed.
DLPX-73059	The issue with the Masking Engine throwing the 'Unsupported Property Error' in application logs for properties that differ in the case from the actual properties' has now been resolved.

Bug Number	Description
DLPX-74740	Masking File Format Import error now shows the list of invalid special characters present in the file name.
DLPX-74760	The issue with the failure of the POST /import with "Unknown document version UNRECOGNIZED" when the source engine version is newer than the destination engine version has now been resolved.
DLPX-75441	The issue with maskedObjectName not populating the execution events when masking files have now been resolved.
DLPX-75487	The issue with DMS_ROW_ID as a column name in the Masking Rule Set causing jobs to fail has now been fixed.
DLPX-75712	The "About" page now lists the correct patent number.
DLPX-75868	The issue with the DataLevel Profiling resulting in an abort with "TypeError: Cannot find function getInteger in object false" has now been resolved.
DLPX-76009	The issue with the failure of the 'File format id greater than a specific number' when trying to update the file format ruleset via the API only has now been resolved.
DLPX-76063	The issue with the failure of the DateShift Algorithm when masking the VSAM (Mainframe) numeric data type has now been resolved.
DLPX-76068	Masking now allows passwords that are longer than 12 characters.
DLPX-76134	The issue with the Welcome screen displaying "User can launch 'Create Job' wizard" when they are not able to have now been resolved.
DLPX-76352	Delimited File masking no longer truncates white-space only fields.
DLPX-76405	Multi-column algorithms now display a better error message when logical fields are missing.
DLPX-76428	For masking operation, the Advanced Oracle Connector now rounds decimal numbers to integers.
DLPX-76450	The Payment Card framework UI now permits configuring minimumMaskedPositions to 0.

Bug Number	Description
DLPX-76493	The issue with the MSSQL instance name property not being passed by default when connecting has now been resolved.
DLPX-76541	The issue with the file masking job failure using a pattern with a Windows-based FTP server has now been resolved.
DLPX-76566	The issue with the profiling Job failure with the 'Couldn't get row from result set' error due to conversion unsupported has now been resolved.
DLPX-76608	The plugin's authorization to delete files in the temp directory is now granted.
DLPX-76610	The issue with the IP SFTP Masking failure to delete the file has now been resolved.
DLPX-76670	The issue with the masking Job failure with the 'Conversion failed from string to uniqueidentifier data type' error has now been resolved.
DLPX-76821	The issue with the throwing of JSchException for pattern-based SFTP masking with file count > 10 has now been resolved.

Release 6.0.9.0

Bug Number	Description
DLPX-57961	Inventory export fails silently when a dataFile has fileFormats = NULL.
DLPX-64329	v5 API: Create an endpoint to copy environment objects in the same/different environment.
DLPX-68807	DateShift algorithm example should exclude invalid entries in the UI pop-up.
DLPX-69728	The active CIFS/NFS mount is getting disconnected after the upgrade.
DLPX-72383	Masking job hangs due to "Unable to acquire lock for job removal before timeout."
DLPX-73344	Internal server error when importing invalid delimited or fixed-width file format.
DLPX-74409	Masking Engine: Upgrade slf4j-ext-1.7.25.jar to slf4j-ext-1.7.30.jar.
DLPX-74415	Masking Engine: Upgrade Guava version to 30.1-jre.

Bug Number	Description
DLPX-74882	Masking's SFTP client no longer compatible with SolarWinds and Goanyware SFTP servers.
DLPX-74913	Inventory exports do not include the notes field.
DLPX-74941	Create a sync state on export for syncable objects that have null sync states.
DLPX-75005	Importing the COMPONENT type algorithm does not change the sync state object type.
DLPX-75202	Batch Masking and Failed kettle jobs may fail to terminate.
DLPX-75235	Secure lookup GUI: Add support to specify remote file URI.
DLPX-75244	Extensible driver test fails "Parameter 'directory' is not a directory" for the removed driver.
DLPX-75296	Sync import fails for an object having files with space in the filename.
DLPX-75307	Multi-column algorithm assignment details are missing from CSV inventory export.
DLPX-75308	SQLFeatureNotSupportedException method not supported ...getSchema().
DLPX-75311	Debug message with %s logged when using Regex Decomposition Algorithm.
DLPX-75437	LastNameSeparator text box is not disabled for default dlpx-core:FullName algorithm.
DLPX-75440	XML masking job fails with "Sequencer step still had unwritten rows!".
DLPX-75468	Upgrade MySQL driver org.mariadb.jdbc:mariadb-java-client from 2.4.1 to latest available version 2.7.2.
DLPX-75516	Updated Masking Web API version to 5.1.9.
DLPX-75520	Fixed an issue that could cause XML masking jobs to stall or fail with the error "Sequencer step still had unwritten rows!".
DLPX-75644	Added new "UserDirIsRoot" flag to the SFTP type connector.
DLPX-75768	Row limiter can still deadlock jobs in some failure cases.

Bug Number	Description
DLPX-76267	Sync export fails with insufficient memory available in JVM error.

Release 6.0.8.0

Bug Number	Description
DLPX-66147	Environment errors occur after deleting a referenced Mainframe connector.
DLPX-71318	Transformation - SQL check for CREATE and DROP IDENTITY Column is not using Schema.
DLPX-71489	Masking plugin API does not include the plugin author from Jar metadata.
DLPX-72581	Masking usernames and emails not redacted in support bundles.
DLPX-72653	Masking Job "Row Limit" UI shows 20 to be the lowest limit - This has been fixed to reflect 100 as the lowest.
DLPX-73207	Table name for MSSQL with single quote appears incorrectly on inventory page.
DLPX-73328	Incorrect tooltip text displayed for Admin link in footer.
DLPX-74152	Unable to edit ruleset from UI after adding tab (4 space) as an "End Of Record" in file ruleset.
DLPX-74190	Sync import of global settings fails with NullPointerException in an extended algorithms tearDown method.
DLPX-74426	PostgreSQL driver got updated from 42.2.10 to 42.2.19 version.
DLPX-74612	Oracle Masking Job fail with FanManager - unable to create ONS subscriber.
DLPX-74638	Bad example format in Date Algorithm GUI.
DLPX-74844	Algorithm UI breaks with JSON special characters in the algorithm extension JSON.
DLPX-74849	Adding a new field to a record type via the GUI incorrectly always sets the field to be masked.

Bug Number	Description
DLPX-74875	Importing pre/post script into the same environment with the same file name and job name deletes the file.
DLPX-74881	Certain algorithm plugins causes minor breakage in Algorithm Settings Screen.
DLPX-74967	New Date Shift algorithms do not allow for any time zone specifiers in the date format.
DLPX-74974	InvalidKeyException "No installed provider supports this key: (null)".
DLPX-74990	Specifying Backspace character("\b") as enclosure for delimited files via API does not throw an error, but crashes UI.
DLPX-75246	Mask Value Range for Segment Mapping (legacy) not getting saved from GUI.
DLPX-75290	Cannot use MSSQL or JTDS driver in SDK as extensible framework.

Release 6.0.7.0

Bug Number	Description
DLPX-45399	Improve masking test connector errors.
DLPX-57910	Control character field delimiters are replaced incorrectly in delimited file masking.
DLPX-67246	The UI and the API should have the possibility to LOCK a user account.
DLPX-70837	Update MDS "All Privileges" role to have correct privileges.
DLPX-70844	End of Record options for file masking is misleading.
DLPX-70885	Masking API to submit update password request with forgot password token.
DLPX-71125	Masking Bundle generation is very slow.
DLPX-72036	UI sync operations initiate but fail; no evidence in MDS or logs.
DLPX-72121	Algorithm description field limit on UI should be same as new API limit i.e 8192.

Bug Number	Description
DLPX-72424	String masking algorithm results in null values when masking oracle LONG(0) columns.
DLPX-72501	Regression in delimited file allowed Delimiters.
DLPX-72509	DateShift cast of DATE to DATETIME is not range cognizant.
DLPX-72551	FreeTextRedactionExtension translator does not properly set profileSetId when API version is v5.1.3 or less.
DLPX-72731	Incorrect handling end-of-record (EOR) character embedded in an enclosure.
DLPX-72734	The plugin VIEW privilege is no longer required to add, update, or delete a plugin.
DLPX-72878	Migration V2019.04.11.0 wrongly assumes role with role_id==1 always present.
DLPX-72879	Extensible algorithm numeric to string conversion is inconsistently producing input String with scientific notation.
DLPX-73068	Fixed an issue that causes numeric algorithms using the extensibility framework to fail when applied to fixed-width files.
DLPX-73157	Masking job queued failing immediately as unable to get the execution ID.
DLPX-73187	Custom sql inside the ruleset is not getting auto-generated in case the custom property file is used.
DLPX-73302	Remove GUI validation to support multiple characters for the delimiter.
DLPX-73327	Job with multiple tables/files that differs only by case run indefinitely.
DLPX-73384	Special characters in mysql database instance names are not properly escaped.
DLPX-73441	Masking IP on DB2 using 'Direct Row Access' with ROWID is failing with conversion error.
DLPX-73477	Prevent locked user accounts from logging in when SSO is enabled.
DLPX-73599	Fixed an issue that causes loss of sub-millisecond precision when processing MS SQL Server datetime types.

Bug Number	Description
DLPX-73671	Uploading Hive driver on the masking engine is failing with InsufficientJvmPermissionException.
DLPX-73702	Extended Connector Profile Job fails with FilePermission required for "target": "/tmp/jtids2094637632459524041.tmp" with "action": "write".
DLPX-73805	Masking UI: SM editor spins when create 4 * alpha-numeric segments.
DLPX-73886	Upgrade Masking API version to v5.1.7.
DLPX-74055	Allow masking admin users to have api access rights revoked.
DLPX-74135	Empty string delimited inside of enclosures results in masking job failure.
DLPX-74185	Character Mapping algorithms with more than 3 characterGroups do not display correctly in UI.
DLPX-74188	Masking connector properties API/UI needs to redact passwords.
DLPX-74292	Custom property file is getting ignored for the source connector in case of OTF job resulting in job failure.

Release 6.0.6.0

Bug Number	Description
DLPX-59842	Fixed an issue causing jobs to fail with out of memory or stack overflow exceptions when the number of tables exceeded a threshold of approximately 800 per stream. It should no longer be necessary to set job streams greater than 1 to avoid this issue.
DLPX-64493	The Roles API is missing elements for the following categories: Custom Algorithms, Diagnostic, Inventory Report, and Approve Inventories.
DLPX-71396	Settings link is missing from footer for user without setting permissions
DLPX-71397	Settings link in footer redirects to profilerSettings.do instead of default jdbcDriver.do
DLPX-71830	Database Tokenize/re-identify job's commit size is not set to default post-upgrade

Bug Number	Description
DLPX-72079	MSSQL JDBC Urls should accept 'database' as a valid parameter
DLPX-72095	Some extended connectors db drivers - throw errors for connection properties they don't understand
DLPX-72311	Exposed DEFAULT_MULTIPHI_ALGORITHM setting via API.
DLPX-72385	Edit Custom Algorithm - Name of Previously Uploaded File No longer Visible.
DLPX-72460	Large environment export hangs.
DLPX-72564	"Add Application" option should be on top inside the action dropdown list.
DLPX-72704	Expanded LK table text limit 1024 characters.
DLPX-72867	Mssql driver is not working with the extended connector in case the instanceName is given in the JDBC url.
DLPX-73082	Unable to assign algorithm to XML fields which contain special characters.
DLPX-73212	Copying an environment that contains a profile or tokenization job causes the environment export to fail with NullPointerException.
DLPX-73338	XSS attack is getting executed on the environment overview page.
DLPX-73502	OTF job with generic connector is failing.

Release 6.0.5.0

Bug Number	Description
DLPX-62372	API authorization token used by the UI expires before the UI login session.
DLPX-70685	Removal of format installation via FTP, SFTP, and mount for XML and Mainframe File Format.
DLPX-71387	Editing recordType to change recordTypeQualifier results in empty JSON.

Bug Number	Description
DLPX-71540	Added Application option is not displayed to the user without copy environment permission.
DLPX-71686	Deleting all mountFilesystem objects nor rebooting does not stop the running portmapper and auxiliary NFS RPC services.
DLPX-50282	Masking support for Oracle XMLType.
DLPX-71666	Characters in Ignore Characters causes Non-Conforming error in Segment Mapping.
DLPX-71758	Propagated SSL related system properties set in Tomcat to Kettle.
DLPX-71734	Masking SQL Server datatype datetime2 generate conversion error.
DLPX-71824	DB-To-File masking job failure.
DLPX-71159	Uploading copybook file format fails if a filename contains multiple full stops.
DLPX-71915	Segment mapping doesn't mask and reports success when positions are misconfigured.
DLPX-71531	Extended algorithm internal conversion of numeric to string types produces unexpected results.
DLPX-72003	Newline characters in the description of an extended algorithm break the Algorithm Settings UI.
DLPX-72028	Using Algm-SDK 1.1 on Windows, algm builds fail w/ 'Illegal char <:> at index 2:'.
DLPX-72128	Overly aggressive quoting of Oracle usernames breaks proxy users.
DLPX-72194	Upgraded MSSQL driver to latest version 8.4.1.
DLPX-72267	Made default API version configurable through application settings.
DLPX-72263	Domain value is not retained on defining a file field causing NPE while job execution.
DLPX-72308	RPC serviceUser can delete an active mount which resulted in active RPC services.

Bug Number	Description
DLPX-72367	Null Pointer Exception when applying a String type extended algorithm or non-legacy Secure Lookup to numeric type columns.

Release 6.0.4.0

Bug Number	Description
DLPX-69407	Hybrid jobs are not syncable.
DLPX-69476	File connector sync throws an error for missing passwords.
DLPX-69834	The user without permission can access UI components using a direct URL.
DLPX-70053	VSAM job performance still poor when file wildcards are used due to flaw in DLPX-68780 fix.
DLPX-70265	NPE along with 'problem-saving mapplet' pop-up is displayed for invalid filereferenceld.
DLPX-70412	OTF Masking SYBASE could not mask 2 tables with the same name but different owners.
DLPX-67886	Updated the SAP ASE (Sybase) JDBC Driver.
DLPX-70567	Implemented a job queue to regulate memory consumption.
DLPX-70642	Copy Ruleset performance improvement.
DLPX-69699	VSAM Masking - Inventory blank after Copy Rule Set fails to copy and corrupts Rule Set and File Format.
DLPX-67501	Fixed an issue that caused Delimited and Fixed-width data level profiling jobs with an FTP or SFTP connector to hang on large files.
DLPX-63065	Updated jquery.js library for Masking to 1.12.0d.
DLPX-69124	Fixed an issue discovering column metadata for Oracle databases that could result in incorrect column lengths and masking jobs failing on update because values are not trimmed correctly.

Bug Number	Description
DLPX-70651	application_nm is not trimmed automatically during an upgrade.
DLPX-70878	Fixed an issue where an on-the-fly Masking job with the disable constraints feature on attempted to use null as the database password.
DLPX-63491	File Masking OTF jobs create the file at the end of the job instead of continuously writing masked rows.
DLPX-59952	OutOfMemory in File Masking when masking large or many files.
DLPX-70395	Renamed Delphix FT algorithm properties "Blacklist" and "Whitelist" to "Denylist" and "Allowlist".
DLPX-70807	Removed Row Types for Database Inventory.
DLPX-70662	Removed Scheduler from Masking.
DLPX-71000	Fixed an issue where CLOB and NCLOB masked values were being incorrectly truncated on Oracle. Refresh the ruleset for the fix to take effect.
DLPX-70982	Masking LDAP user is locked locally when LDAP auth fails.
DLPX-71235	In the monitor screen, all tables show failed if any tables are failed.
DLPX-71320	Removed/hid the environment export checkbox from the roles page.
DLPX-71310	The profiling job fails if a profiler set matches all columns of a table using column profiling.
DLPX-71424	Disable triggers, drop constraints, drop indexes, prescripts and postscripts target source database with OTF jobs and advanced connectors.
DLPX-71530	Unmasked values with only spaces result in (null) masked value.

Release 6.0.3.0

Bug Number	Description
DLPX-63874	ExecutionComponent status for unwritable files was incorrect when masking over SFTP.
DLPX-68123	Masking Engine does not re-read Kerberos config dynamically.
DLPX-68725	Upgraded tomcat to 9.0.31 or later.
DLPX-69655	loginid did not support '@' when creating connectors.
DLPX-69492	MSSQL driver requires java.net.socketpermission to accept permission which is not present in MDS.
DLPX-69493	Execution event is not getting generated for profile job in case of missing permission.
DLPX-69761	Masking Jobs, fail to save added Pre-Scripts.
DLPX-69766	Masking GUI: Remove any script from masking job dialog removes both the scripts.
DLPX-69782	Export/Import Environment using engine sync API.
DLPX-69780	UI based Export Global Object using engine sync API.
DLPX-46853	Switch from jTDS to Microsoft SQL Server JDBC driver.
DLPX-65380	Masking Jobs with commit size >= 340 are getting failed on Azure Managed SQL instance.
DLPX-69815	Secure_shuffle algorithm fails for decimal data type using extended connector.
DLPX-69806	Inventory UI is susceptible to URL based XSS attack.
DLPX-69779	Mapplet's input and output fields are susceptible to XSS attack.
DLPX-69832	Import Environment using sync API.
DLPX-69833	UI: Import Global Object using sync API.
DLPX-69861	Define Fields 'Field Name' input is susceptible to XSS attack.

Bug Number	Description
DLPX-69888	XSS script in file pattern is getting executed.
DLPX-69960	Unable to Edit File format if the Enclosure is set to " (double quote).
DLPX-69671	Delimited File Masking with delimiter inside enclosure is handled incorrectly.
DLPX-69922	Inventory UI is susceptible to XSS attack using malicious column names.
DLPX-69941	Error report on job monitor page is susceptible to XSS attack.
DLPX-69989	dateFormat field of date algorithms is susceptible to XSS attack.
DLPX-69920	Import/Upload file UI is susceptible to iframe based XSS attack, throughout the application.
DLPX-69919	Redaction value input field of Free Text Redaction algorithm is vulnerable to XSS attack.
DLPX-69917	Export Inventory UI is susceptible to URL based XSS attack.
DLPX-70055	Masking - Inventory for oracle always picking up NUMBER (22) instead of real NUMBER definition.
DLPX-70046	OTF job with decimal data type and secure shuffle algorithm is changing the last digit after the decimal point of the unmasked column in case of Hana database.
DLPX-70050	CSV and XML file masking performance improvements.
DLPX-70074	Copying an environment does not create a sync state.
DLPX-69851	Masking jobs fail to set fetch size large enough in the input step query.
DLPX-69672	Delimited File Masking and Segment Mapping is not ignoring delimiter if specified as ignore character.
DLPX-69954	Delimited file masking row parsing incorrect when a field contains multiple enclosure characters and a delimiter.
DLPX-70178	Delimited Files: Improve validation for delimiter and enclosure from API.

Bug Number	Description
DLPX-70182	Improved validation for delimiter and enclosure from GUI.
DLPX-70217	"Max number of jobs" Setting on masking engine should be API accessible.
DLPX-70379	For the multi-tenant job, the source connector dropdown doesn't show the connector in the list if the connector instance name contains the space in between.
DLPX-70558	searchEnvironment parameter in URL is vulnerable to XSS attack.
DLPX-70557	Copy Ruleset has a scale performance issue with a large number of tables/columns.
DLPX-70641	Unmasked data logged in the support bundle logs when using extended connector with enable_logger functionality on

Release 6.0.2.0

Bug Number	Description
DLPX-65833	Removed unnecessary error out on passwords being provided for file connectors using the mount mode.
DLPX-65319	New API endpoint for mainframe-dataset-record-type.
DLPX-68153	If creating a mapping algorithm in the Masking UI fails, the failure is now properly reported to the user.
DLPX-67882	Upgrade the PostgreSQL JDBC driver to version 42.2.10.
DLPX-58184	List rule sets alphabetically on the inventory page.
ES-662	Added Sync support for data set connectors.
ES-664	Added Sync support for mainframe data set formats
ES-671	Added Sync support for Mainframe data set jobs
ES-665	Added Sync support for Mainframe data set rule sets.
DLPX-68786	Masking job misreported successful tables as 0 rows masked.

Bug Number	Description
DLPX-67517	Added support for on-the-fly jobs from a database to a delimited file.
DLPX-68842	Jobs slowed down over time - after running many jobs.
DLPX-68985	A memory leak occurred for Informix/oracle database on every test connection using an extended connector.
DLPX-68780	VSAM Input step performance was negatively affected by the number of unmasked fields.
DLPX-67886	Sybase jConnect driver failed when a batch contains string parameters of different sizes and HOMEGENOUS_BATCH=true.
DLPX-65841	Fixed an issue where a REST API call to GET /syncable-objects? object_type=MASKING_JOB would fail after environment copy.
DLPX-69156	Test Connection always returned connection succeeded in case of wrong jdbc url with extended connector.
DLPX-69238	Secure Shuffle algorithm, when used with extended connectors, left data unmasked but reports success.
DLPX-69244	Importing a 5.3.x Masking Environments into 6.0.1 ME, the Application Name is converted to numeric.
DLPX-69154	Fixed an issue where setup could fail if the DNS Domain is empty.
DLPX-69622	Data level profiling jobs fail with "Couldn't find field 'XYZ' in row!"

Release 6.0.1.0

Bug Number	Description
DLPX-64530	Allow a JDBC URL to contain a single quote (') character.
DLPX-65302	Add a status column to the audit log page to report each recorded action's result (success/failure).
DLPX-65622	Fix an issue where an in-place, multi-tenant XML file masking job that used file patterns did not have an execution component.

Bug Number	Description
DLPX-65974	Updated log statements in the file masking job logs to reflect that file connectors may use mounts in addition to FTP and SFTP.
DLPX-66127	Fixed a job monitoring issue when counting the rows in table with more than 2+ billion (2,147,483,647) rows.
DLPX-62130	Fixed an issue with the XML file inventory GUI that prevented users from assigning algorithms to both a tag and its attribute(s).
DLPX-66272	Fixed an issue where an on-the-fly job using generic connectors used an incorrect database password.
DLPX-66600	Removed the requirement to restart the Masking service after changing email settings.
DLPX-66328	Fixed an issue with file masking jobs using multiple record types that could cause the job to fail or corrupt the output.
DLPX-66557	Added support to the Date Shift algorithm for numeric data types.
DLPX-66517	Enhanced the GET /file-field-metadata endpoint to return the full XML XPath for an XML field.
DLPX-66102	"Drop Indexes" checkbox now handles compound indices correctly for Sybase.
DLPX-66967	Fixed a Job Scheduler issue that caused a periodic job to only running once.
DLPX-67318	Prevent reordering of the XML file inventory GUI when an algorithm is assigned
DLPX-67317	On the XML file inventory GUI, open the algorithm assignment dialogue box with a single mouse click
DLPX-66076	Added API endpoints for file recordTypes and recordTypeQualifiers
DLPX-65855	Optimize the performance of EngineSync import, export, and get syncable object for large database rule sets.
DLPX-65987	Fixed an issue that caused data level profiling of a database to fail when a column name was a special JavaScript word.

Bug Number	Description
DLPX-67747	Fixed an issue that caused some delimited or fixed file masking jobs with multiple record types of different lengths to fail.
DLPX-67470	Fixed delimited file masking to treat double quote (") characters in fields as normal characters.
DLPX-67765	Updated the Sybase JDBC driver.
DLPX-67838	Fixed an issue that prevented XML File masking jobs from scaling above a few thousand files.
DLPX-67832	Non-administrators can no longer regenerate the engine encryption key.
DLPX-67960	Make username searches on the Audit page case insensitive.
DLPX-68148	Fix an issue that caused an XML file masking job to run out of memory when masking very large XML input files.
DLPX-46220	Import of extremely large object sets via the GUI XML feature is handled inefficiently.

Release 6.0.0.0

Bug Number	Description
DLPX-42385	Added a job execution event with information on how to resolve an Oracle deadlock error (ORA-00060), see https://www.delphix.com/masking-help/knowledge-base/KBA1853 .
DLPX-47004	Added a job execution event with information on how to resolve an Oracle snapshot too old error (ORA-01555), see https://www.delphix.com/masking-help/knowledge-base/KBA1827 .
DLPX-47662	Test connector detects that a file/mainframe connector targets a single file instead of a directory and fails.
DLPX-52151	Fixed copy rule set to prevent leading/trailing spaces in a new rule set's name.
DLPX-55478	Correctly display file patterns, including escape characters, throughout the user interface.

Bug Number	Description
DLPX-55739	Fixed the disable constraint feature to support an Oracle constraint (a) created by a different database user than the Masking job's database user and (b) using a validation setting of "NOT VALIDATED".
DLPX-58958	Added support for LDAPS (LDAP over TLS/SSL).
DLPX-59060	Attach the correct PDF report to all job execution emails.
DLPX-59111	When editing a large rule set in the GUI, do not reset to the first page after editing and saving a modification to a rule set component.
DLPX-59807	If a failure occurs during job generation, do not attempt to execute the job.
DLPX-60200	When uploading an SSH key, return an error if the name contains one of the following restricted characters: \ (backslash), ; (semi-colon), % (percent), ? (question mark), or : (colon).
DLPX-61630	Improved the performance for appending new mapping values to a mapping algorithm.
DLPX-62214	Fixed PDF report download URLs.
DLPX-62593	Fixed creation of a PDF audit report on the Audit tab of the user interface.
DLPX-63365	Removed leading/trailing spaces from Masking object names on upgrade. For naming rules, see the Getting Started > Naming Requirements section in the documentation.
DLPX-63706	Fixed the XML file inventory GUI to show an algorithm edit button for a tag with the same name as its parent.
DLPX-64691	Added support in the user interface for Cobol copybooks with a redefine condition at level 01.
DLPX-64707	Improved the file record types user interface to (a) remove the unnecessary length input and (b) clarify that the qualifier may be a regular expression.
DLPX-65274	Improved the performance of the copy environment feature.
DLPX-65314	Fixed an issue in the copy environment feature that removed file format assignments from the source environment.

Bug Number	Description
DLPX-65632	Fixed an issue in the segment mapping algorithm that caused duplicate mappings if a minimum value was specified for the real values range.
DLPX-65860	For mainframe file masking, add support for a redefine condition on a field name that contains a - (dash) followed by a digit.
DLPX-65866	Fixed an issue with the rule set GUI when displaying table names longer than 50 multi-byte characters.

Known issues

Release 13.0.0.0

Key	Summary	Workaround
DLPX-85469	JSON file masking does not support the use of multi-column algorithms on (a) Fields in two or more different arrays (b) Fields at different levels in a single multi-dimensional array.	None

Release 12.0.0.0

Key	Summary	Workaround
DLPX-85469	JSON file masking does not support the use of a multi-column algorithms on <ul style="list-style-type: none"> Fields in two or more different arrays. Fields at different levels in a single multi-dimensional array. 	No workaround.

Release 11.0.0.0

Key	Summary	Workaround
DLPX-86196	Segment Mapping algorithms exported using Engine Synchronization from releases 6.0.15.0 to 10.0.0.0 are not validated upon import to 11.0.0.0.	View the imported algorithm configuration in the UI and click save.
DLPX-84875	User with create/edit permissions is not able to create or edit the classifier	Assign domain view role to user.
DLPX-85469	JSON file masking does not support the use of a multi-column algorithms on <ul style="list-style-type: none"> Fields in two or more different arrays. Fields at different levels in a single multi-dimensional array. 	No workaround.
DLPX-85560	Salesforce on the fly jobs fail with latest CDATA driver, with a license error.	No workaround.

Release 10.0.0.0

Key	Summary	Workaround
DLPX-85469	JSON file masking does not support the use of a multi-column algorithms on <ul style="list-style-type: none"> Fields in two or more different arrays. Fields at different levels in a single multi-dimensional array. 	No workaround.
DLPX-85560	Salesforce on the fly jobs fail with latest CDATA driver, with a license error.	No workaround.
DLPX-84875	Unable to create/edit the classifier/expression using the user with create/edit permission of profiler role	Assign domain view role to user.
DLPX-85867	Incorrect non-conforming data count across all the tables of a job, if at least one table has non-conforming data.	No workaround.

Release 9.0.0.0

Key	Summary	Workaround
DLPX-84875	Unable to create or edit classifiers/expressions as a user with create/edit permission of the Profiler role.	Assign the domain view role to user.
DLPX-85867	Incorrect non-conforming data count across all the tables of a job, if at least one table has non-conforming data.	No workaround.

Release 8.0.0.0

Key	Summary	Workaround
DLPX-84141	Masking Engine environment revisionHash changes on profile job execution even though there is no change in inventory	No workaround.

Key	Summary	Workaround
DLPX-85867	Incorrect non-conforming data count across all the tables of a job, if at least one table has non-conforming data.	No workaround.

Release 7.0.0.0

Key	Summary	Workaround
DLPX-84141	Continuous Compliance Engine environment revisionHash changes on every profile job execution even if there is no inventory change.	No workaround.
DLPX-84525	Inventory Page fails to load on Continuous Compliance Engines when there are more than 32,767 masked columns.	A possible workaround is to reduce the number of masked columns on the Engine to less than 32,765.
DLPX-85867	Incorrect non-conforming data count across all the tables of a job, if at least one table has non-conforming data.	No workaround.

Release 6.0.17.0

No known issues in this release.

Release 6.0.16.0

Bug Number	Description	Workaround
DLPX-82517	Issues w/ DB2 iSeries Connector License Installation	No workaround.

Release 6.0.15.0

Bug Number	Description	Workaround
DLPX-81895	Data Profiling results are not shown at Job -> Monitor page -> Results tab, in the case of delimited, fixed, and XML files.	No workaround.

Bug Number	Description	Workaround
DLPX-82517	Issues w/ DB2 iSeries Connector License Installation	No workaround.

Release 6.0.14.0

Bug Number	Description	Workaround
DLPX-82517	Issues w/ DB2 iSeries Connector License Installation	No workaround.

Release 6.0.13.0

No known issues in this release.

Release 6.0.12.0

Bug Number	Description	Workaround
DLPX-78478	Reidentification of CM numeric algorithm on decimal data is failing.	When the field is long enough, use the Tokenization algorithm instead of CM tokenization.
DLPX-78659	CM Numeric is not producing unique results for the floating-point numbers.	Use an algorithm other than CM Numeric algorithm for masking floating-point numbers stored in a numeric field.
DLPX-79567	Drop index fails if an index with the same name exists on the masked columns across multiple tables for MSSQL databases.	No workaround.
DLPX-79803	Masking with MSSQL database fails if table name contains '['.	No workaround.
DLPX-79804	Masking with MSSQL database fails if table name contains '\'.	No workaround.

Release 6.0.11.0

Bug Number	Description	Workaround
DLPX-78009	The masking job fails when masking the primary key column if Drop Indexes are not enabled along with the Disable Constraints.	In addition to Drop Indexes, you must enable Disable Constraints when masking primary keys using built-in driver support functionality. Advanced users who are not satisfied by some limitations of the built-in Oracle support for masking primary keys may also create custom pre and post-scripts to perform both drop indexes and disable constraints operations.
DLPX-79803	Masking with MSSQL database fails if table name contains '['.	No workaround.
DLPX-79804	Masking with MSSQL database fails if table name contains '\'	No workaround.

Release 6.0.10.0

No known issues in this release.

Release 6.0.9.0

No known issues in this release.

Release 6.0.8.0

Bug Number	Description	Workaround
DLPX-74882	Masking's SFTP client no longer compatible with SolarWinds and Goanyware SFTP servers	No workaround.

Release 6.0.7.0

No known issues in this release.

Release 6.0.6.0

No known issues in this release.

Release 6.0.5.0

No known issues in this release.

Release 6.0.4.0

No known issues in this release.

Release 6.0.3.0

No known issues in this release.

Release 6.0.2.0

Bug Number	Description	Workaround
DLPX-69638	Masking job created on engine 6.0.1.1 or prior is failing after the upgrade to version 6.0.2.0 or later	Masking jobs created in 6.0.1.x using a HANA JDBC driver will need to be updated to grant the following permission

Release 6.0.1.0

No known issues in this release.

Release 6.0.0.0

Bug Number	Description	Workaround
DLPX-60397	If a mapping algorithm is included in multiple jobs, only one job should be run at a time. If multiple jobs are run at the same time, then the mapping algorithm might contain multiple mappings to the same value or the jobs might deadlock.	Only run one job at a time.
DLPX-61405	Masking operation should wait for zfs delete queue to drain	Replication may send more data than expected if masking involves dropping large DBF files.
DLPX-74882	Masking's SFTP client no longer compatible with SolarWinds and Goanyware SFTP servers	No workaround.
DLPX-64493	V5 API /roles endpoint missing certain items	View and set these privileges through the GUI

Bug Number	Description	Workaround
DLPX-66973	Date format is changed after importing the environment	Either (a) use the GUI import feature and then review the imported date formats for correctness or (b) use EngineSync to export/import jobs, which will not alter the date format.

Deprecated and end-of-life features

Release 11.0.0.0

Deprecated features

- **Job wizard UI page**

The job wizard is deprecated as of this release. The job wizard offers a simplified job creation experience, but only for a limited set of data platforms. We are incrementally upgrading the existing Continuous Compliance UI pages with new, reactive interfaces and improved user experiences. As part of that process, the job wizard's capabilities will be incorporated into the product's standard workflows, which makes them available for all data sources.

End-of-Life features

- **ESXi 6.0 support**

Release 10.0.0.0

Deprecated features

- **ESXi 6.5 and 6.7** VMware's end of technical guidance for ESXi is on 11/15/2023. To align our support, Delphix will end support of ESXi 6.5 and 6.7 in version 15.0.0.0 (November 2023 release).

Release 6.0.17.0

End-of-life features

- **Oracle 11.1 and 12.1** Details of the Oracle database end-of-life can be found in the [Oracle Lifetime Support Policy](#).

Release 6.0.15.0

End-of-life features

- **Legacy algorithms & mapplets:** The last algorithm migrations have completed in this version (6.0.15), Legacy Algorithms are now EOL. For more information, see this [Delphix Community Post](#).

Release 6.0.12.0

End-of-life features

- **Legacy Secure Lookup:** Legacy Secure Lookup has been removed and only the extensible version is supported. Previous secure lookup instances have been moved to the extensibility framework.
- **Internet Explorer 11 support:** Internet Explorer 11 is no longer supported by Delphix. Users are requested to refer to the list of [Supported browsers](#).

Release 6.0.11.0

Deprecated features

- **Oracle 11.1 and 12.1:** Details of the Oracle database end-of-life can be found in the [Oracle Lifetime Support Policy](#).
- **TLS 1.0 and 1.1** These versions of TLS are known to be vulnerable, enterprise use is heavily discouraged.

End-of-life features

- **Oracle 10g support:** 6.0.10.0 is the last release supporting Oracle 10.1 and 10.2.
- **Create/update of legacy secure lookup algorithms via UI:** The ability to create and update legacy secure lookup algorithms has been removed from the UI. This feature is still accessible through the API endpoints.

Release 6.0.10.0

End-of-life features

- **Ruleset Edit:** The Table Suffix, Add Column, Join Table, and List options were deprecated in the 6.0.3.0 release. These options have reached the end of life in the 6.0.10.0 release and have been completely removed from the product. These options are the rarely used feature that can be achieved using the following alternatives:
 - If you were using the Table Suffix functionality, you can achieve the same results with a series of API calls (/table-metadata and /column-metadata endpoints).
 - For Add Column, Join Table, and List, you need to convert these settings to the equivalent Custom SQL configuration before upgrading to 6.0.10.0 release.

Release 6.0.9.0

Deprecated features

Delphix has been creating new and improved versions of our existing algorithms, thus, Delphix would like to provide formal notice of deprecation and planned End-of-Life (EoL) for the older algorithm versions. This is to inform our customers that planning should start for their transition to these updated algorithms. For more information and details on the transition, see [Delphix Community Post - Legacy Mapping Algorithm](#).

Release 6.0.8.0

End-of-life features

- Legacy Custom Algorithm (Mapplet). For more information, see [Delphix Community Post - Mapplet EoL](#).
- SAP ASE (Sybase) 15.0.3 support

Release 6.0.7.0

End-of-life features

- ESX 5.5 support
- Masking Connectors: Db2 LUW and zOS v9, Db2 LUW and zOS v10, SQL Server 2005, 2008, 2008 R2

Release 6.0.4.0

Deprecated features

- **FTP, SFTP, and mount upload for XML and Cobol formats** FTP/SFTP/Mount-based format import were the original modes for XML and Cobol files, since then, Delphix has added the ability to upload a format file, which is simpler to set up. After the introduction of “upload”, there has been a dramatic shift away from the legacy import modes in favor of the simplicity of “upload”.
- **Row type feature** Originally geared for limiting masking to subsets of rows within a column, this feature was seldom used. The functionality, if desired, can still be replicated via the Custom SQL feature.
- **Redundant settings for ‘edit table’ under rule sets** Table Suffix, Add Column, Join Table, and List - These settings are redundant and can be replicated with the Custom SQL setting.
- **‘HAVING’ clause from Masking API** Deprecating due to low use. This feature, if desired, can be replicated with Custom SQL.

End-of-life features

- **Job Scheduler** As of this release, Delphix has removed the Job Scheduler feature. The introduction of Masking’s REST API several releases ago allowed customers to schedule job executions using their preferred job scheduler. As a result, the integrated scheduler is seldom used.

Release 6.0.3.0

End-of-life features

In this release, the deprecated XML import/export functionality has been removed. If you used the XML import/export feature in previous releases, you'll find the new Sync Environment feature to be a more robust and complete solution with complete API support in addition to being available in the UI.

Release 6.0.0.0

End-of-life features

- Native XML CLOB masking: After the upgrade, columns masked as XML CLOBs will have the NULL SL algorithm assigned.
- Excel files can still be masked by first converting them to Delphix-supported file types (CSV, etc). Also, XML CLOBs can be masked by extracting their values into a table (for example - using extractValue in Oracle).
- DB2 9.1, 9.5, and other 9.x versions of LUW & Z/OS
- “Create target” job option: After upgrading jobs using “create target” will be removed.
- “Bulk data” job option: After the upgrade, jobs using “bulk data” will be turned into non-bulk data jobs.
- Native Microsoft Excel Masking: After the upgrade, MS Excel connectors, rulesets, and jobs will be removed.

Licenses and notices

The Delphix Dynamic Data Platform includes licensed, third-party products from the following companies. These products are copyrighted and all rights are reserved by the respective companies:

- Highcharts, © Highsoft

The Delphix Masking engine includes licensed, third-party products from the following companies. These products are copyrighted and all rights are reserved by the respective companies:

- Kendo UI, © Telerik

Starting with 6.0.3.0, the license info is available via a CLI/API on the engine when logged in as a system administrator.

```
engine> cd license
engine license> getLicense
engine license getLicense *> commit
```

Access to the source code of such third party open source components may be permitted or required in certain instances under the applicable open source licenses by sending an email to <mailto:open-source@delphix.com>.

Getting started

This section covers the following topics:

- [Introduction to Delphix Masking](#)
- [Data source support](#)
- [Installation](#)
- [Naming requirements](#)
- [Users and roles](#)
- [Best practices for defining masking roles](#)
- [Audit logs](#)
- [Kerberos configuration](#)
- [Password vault configuration](#)
- [DB2 connector license installation](#)
- [Continuous Compliance Engine icon reference](#)
- [Delphix masking terminology](#)
- [Changing the IP address of the Delphix Engine](#)
- [Stopping and starting the containerized Continuous Compliance Engine](#)
- [Stopping, starting, and restarting the continuous compliance engine](#)
- [Upgrading the Delphix Continuous Compliance Engine](#)
- [Utilization](#)

Introduction to Delphix Masking

Challenge

With data breach incidents regularly making the news and increasing pressure from regulatory bodies and consumers alike, organizations must protect sensitive data across the enterprise. Contending with insider and outsider threats while staying compliant with mandates such as HIPAA, PCI, and GDPR is no easy task—especially as teams simultaneously try to make their organizations more agile.

To tackle the problem of protecting sensitive information, companies are increasingly scrutinizing the tools they've deployed. Instead of reactive perimeter defenses, security-minded organizations must focus on proactively protecting the interior of their systems: their data. Moreover, while mainstay approaches such as encryption may be effective for securing data-in-motion or data resident in hard drives, they are ill-suited for protecting non-production environments for development, testing, and reporting.

Solution

The masking capability of the Delphix DevOps Data Platform represents an automated approach to protecting non-production environments, replacing confidential information such as social security numbers, patient records, and credit card information with fictitious, yet realistic data.

Unlike encryption measures that can be bypassed through schemes to obtain user credentials, masking irreversibly protects data in downstream environments. Consistent masking of data while maintaining referential integrity across heterogeneous data sources enables Delphix masking to provide superior coverage compared to other solutions—all without the need for programming expertise. Moreover, the Delphix DevOps Data Platform seamlessly integrates masking with data delivery capabilities, ensuring the security of sensitive data before it is made available for development and testing, or sent to an offsite data center or the public cloud.

Delphix Masking is a multi-user, browser-based web application that provides complete, secure, and scalable software for your sensitive data discovery, masking, and tokenization needs while meeting enterprise-class infrastructure requirements. The Delphix DevOps Data Platform has several key characteristics to enable your organization to successfully protect sensitive data across the enterprise:

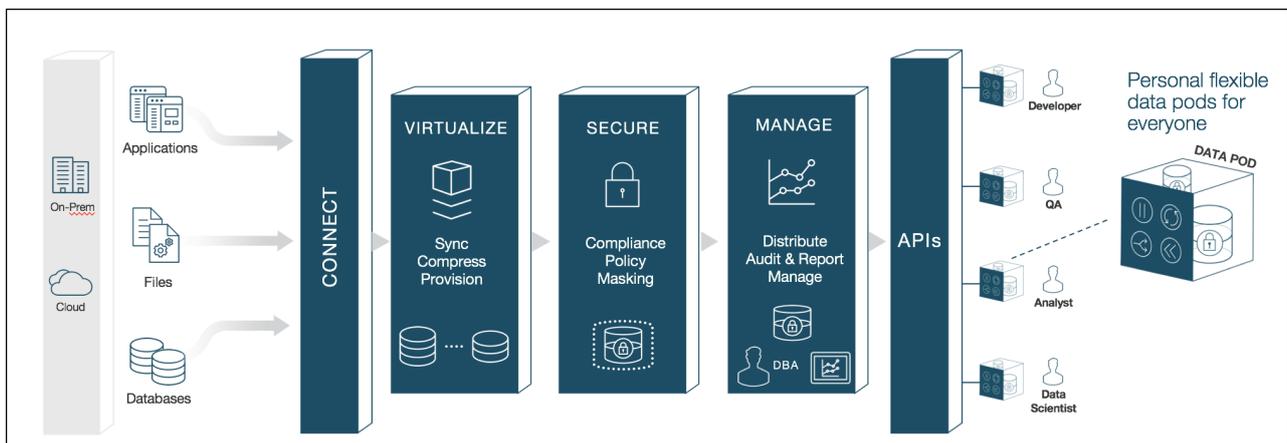
- **End-to-End masking** — The Delphix platform automatically detects confidential information, irreversibly masks data values, then generates reports and email notifications to confirm that all sensitive data has been masked.
- **Realistic data** — Data masked with the Delphix platform is production-like in quality. Masked application data in non-production environments remain fully functional and realistic, enabling the development of higher-quality code.
- **Masking integrated with Virtualization** — Most masking solutions fail due to the need for repeated, lengthy batch jobs for extracting and masking data and a lack of delivery capabilities for downstream environments. The Delphix DevOps Data Platform seamlessly integrates data masking with [data virtualization](#), allowing teams to quickly deliver masked, virtual data copies on-premises or into private, public, and hybrid cloud environments.
- **Referential integrity** — Delphix masks consistently across heterogeneous data sources. To do so, metadata and data are scanned to identify and preserve the primary/foreign key relationships between elements so that data is masked the same way across different tables and databases.
- **Algorithms/Frameworks** — Eighteen algorithm frameworks allow users to create and configure algorithms to match specific security policies. Over twenty-five out-of-the-box, preconfigured algorithms help businesses mask everything from names and addresses to credit card numbers and text fields. Moreover, the Delphix platform includes prepackaged profiling sets for healthcare and financial information, as well as the ability to perform tokenization: a process that can be used to obfuscate data sent for processing, then reversed when the processed data set is returned.

- **Ease of use** — With a single solution, Delphix customers can mask data across a variety of platforms. Moreover, businesses are not required to program their own masking algorithms or rely on extensive administrator involvement. Our web-based UI enables masking with a few mouse clicks and little training.
- **Automated discovery of sensitive data** — The Delphix Profiler automatically identifies sensitive data across databases and files, and the time-consuming work associated with a data masking project is reduced significantly.

High-level platform architecture

The Delphix DevOps Data Platform is made up of 4 main services each of which plays a very important part in delivering fresh secure data to anybody that needs it. These include:

- **Virtualize** — Delphix compresses the data that it gathers, often to one-third or more of the original size. From that compressed data footprint, Delphix virtualizes the data and allows operators to create lightweight, virtual data copies. Virtual copies are fully readable/writable and independent. They can be spun up or torn down in just minutes. And they take up a fraction of the storage space of physical copies -- 10 virtual copies can fit into the space of one physical copy.
- **Identify and secure** — The Delphix platform continuously protects sensitive information with integrated data masking. Masking secures confidential data -- names, email addresses, patient records, SSNs -- by replacing sensitive values with fictitious, yet realistic equivalents. Delphix automatically identifies sensitive values and then applies custom or predefined masking algorithms. By seamlessly integrating data masking and provisioning into a single platform, Delphix ensures that secure data delivery is effortless and repeatable.
- **Manage** — Data operators can now quickly provision secure data copies -- in minutes -- to users in their target environments. The Delphix platform serves as a single point of control to manage those copies. Data operators maintain full control and visibility into downstream environments. They can easily audit, monitor, and report against access and usage.
- **Self-service** — Provides developers, testers, analysts, data scientists, or other users with controls to manipulate data at will. Users can refresh data to reflect the latest state of production, rewind environments to a prior point in time, bookmark data copies for later use, branch data copies to work across multiple releases, or easily share data with other users.



How Delphix identifies sensitive data

Our platform helps you quickly identify your organization's sensitive data. This sensitive data identification is done using two different methods, column-level profiling, and data-level profiling.

Column-level profiling

Column-level profiling uses REGEX expressions to scan the column names (metadata) of the selected data sources. There are several dozen pre-configured profile expressions (like the one below) designed to identify common sensitive data types (SSN, Name, Addresses, etc). You also have the ability to write/import your own profile expressions.

First Name Expression	<code><([A-Z][A-Z0-9]*)\b[^>]*>(.*?)</1></code>
-----------------------	--

Data-level profiling

Data level profiling also uses REGEX expressions, but to scan the actual data instead of the metadata. Similar to column-level profiling, there are several dozen pre-configured expressions (like the one below) and you can write/import your own.

Social Security Number Expression	<code><([A-Z][A-Z0-9]*)\b[^>]*>(.*?)</1></code>
-----------------------------------	--

For both column and data level profiling, when data is identified as sensitive, Delphix recommends/assigns particular algorithms to be used when securing the data. The platform comes with several dozen pre-configured algorithms which are recommended when the profiler finds certain sensitive data.

How Delphix secures your sensitive data

Delphix strives to make available multiple methods for securing your data, depending on your needs. The two secure methods Delphix currently supports are masking (anonymization) and tokenization (pseudonymization).

Masking

Data masking secures your data by replacing values with realistic yet fictitious data. Seven out-of-the-box algorithm frameworks help businesses mask everything from names and social security numbers to images and text fields. Algorithms can also be configured or customized to match specific security policies.

Before Masking	After Masking
Elon Musk	Jeff Bezos

Tokenization

Tokenization uses reversible algorithms so that the data can be returned to its original state. Tokenization is a form of encryption where the actual data – such as names and addresses – are converted into tokens that do not convey any meaning (with regard to appearance and formatting).

Before Tokenizing	After Tokenizing
226-74-3756	asdfkajsfdaja

Data source support

The Continuous Compliance service supports profiling, masking, and tokenizing a variety of different data sources including distributed databases, mainframes, PaaS databases, and files. At a high level, Continuous Compliance breaks up support for data sources into two categories:

- **Delphix connectors:** These are data sources that the Delphix Engine can connect to directly using built-in connectors that have been optimized to perform masking, profiling, and tokenization. Delphix Connectors are available as Standard Connectors and Select Connectors. Standard Connectors are bundled with the masking engine. Select Connectors are an add-on to the Delphix engine and require a separate installation and configuration process.
- **FEML sources:** FEML (File Extract Mask and Load) is a method used to mask and tokenize data sources that do not have dedicated Delphix Connectors. FEML uses existing APIs from data sources to extract the data to a file, masks the file, and then uses APIs to load the masked file back into the database.

Standard connectors

The Delphix Engine has standard masking connectors for the following data sources:

- **Distributed database:** DB2 LUW, Oracle, MS SQL, MySQL, SAP ASE (Sybase), PostgreSQL, MariaDB
- **Mainframe/Midrange:** DB2 Z/OS, DB2 iSeries, Mainframe data sets
- **Files:** Fixed Width, Delimited, XML

For a detailed view of all the versions, features, etc. Delphix supports each data source - see the sections below.

Select connectors

The Delphix Engine has Select masking connectors for the following data sources:

- **Distributed database:** Salesforce, CockroachDB, and SAP HANA 2.0

For a detailed view of all the versions, features, etc. Delphix supports each data source - see the [Select connector support matrix](#) page.

DB2 LUW connector

Introduction

DB2 for Linux, UNIX, and Windows is a database server product developed by IBM. Sometimes called DB2 LUW for brevity, it is part of the DB2 family of database products. DB2 LUW is the "Common Server" product member of the DB2 family, designed to run on the most popular operating systems. By contrast, all other DB2 products are specific to a single platform.

Support matrix

Platforms	Versions	Feature	Availability
Unix	11.1	TLS/SSL	Available
Linux	11.5	Password Vault	Unavailable
Windows		Kerberos	Unavailable

Platforms	Versions	Feature	Availability
		In-place Masking Mode	
		Multi-tenant	Available
		Streams/Threads	Available
		Batch Update	Available
		Drop Indexes	Available
		Disable Trigger	Unavailable
		Disable Constraint	Unavailable
		Identity Column Support	Unavailable
		On-the-fly Masking Mode	
		Restart Ability	Available
		Truncate	Available
		Disable Trigger	Unavailable
		Disable Constant	Unavailable
		Profiling	
		Multi-tenant	Available
		Streams	Available

TLS/SSL setup

1. Add the database's certificate in the setup application using instructions in the **Adding a certificate** section, in the [TrustStore settings](#) article.
2. Restart the Compliance engine.
3. Create a DB2 connector in Continuous Compliance with the relevant parameters. Upload a properties file for the connector with the following:

```
sslConnection = True
```

Oracle connector

Introduction

Oracle Database (commonly referred to as Oracle RDBMS or simply as Oracle) is a multi-model database management system produced and marketed by Oracle Corporation.

Support matrix

Platforms	Versions	Feature	Availability
Unix	11.2	TLS/SSL	Available
Linux	12c	Password Vault	Available
Windows	12cR	Kerberos	Available
AWS RDS	18c	In-place Masking Mode	
OCI DBaaS on Bare Metal	19c	Multi-tenant	Available
OCI DBaaS on VM	21c	Streams/Threads	Available
		Batch Update	Available
		Drop Indexes	Available
		Disable Trigger	Available
		Disable Constraint	Available
		Identity Column Support	Available
		On-the-fly Masking Mode	
		Restart Ability	Available
		Truncate	Available
		Disable Trigger	Available
		Disable Constant	Available
		Profiling	

		Multi-tenant	Available
		Streams	Available

TLS/SSL setup

1. Add the database's certificate in the setup application using instructions in the **Adding a certificate** section, in the [TrustStore settings](#) article.
2. Restart the Compliance engine.
3. Create an Oracle connector in Continuous Compliance with JDBC URL, similar to below examples, depending upon the Oracle database and certificate configuration.

Note: SSL connection with BASIC Oracle connector is not supported.

```
jdbc:oracle:thin:@(DESCRIPTION=(ADDRESS=(PROTOCOL=tcps)(HOST=servername)
(PORT=2484))(CONNECT_DATA=(SERVICE_NAME=servicename)))
jdbc:oracle:thin:@(DESCRIPTION=(ADDRESS=(PROTOCOL=tcps)(HOST=servername)
(PORT=2484))(CONNECT_DATA=(SID=SID_NAME)))
jdbc:oracle:thin:@(DESCRIPTION=(ADDRESS=(PROTOCOL=tcps)(HOST=servername)
(PORT=2484))(CONNECT_DATA=(SID=SID_NAME)
(SEcurity=(SSL_SERVER_CERT_DN="CN=<certificate_cn_name>"))))
```

MS SQL connector

Introduction

Microsoft SQL Server is a relational database management system developed by Microsoft. As a database server, it is a software product with the primary function of storing and retrieving data as requested by other software applications—which may run either on the same computer or on another computer across a network (including the Internet).

Support matrix

Platforms	Versions	Feature	Availability
Unix	2012	TLS/SSL	Available
Linux	2014	Password Vault	Unavailable
Windows	2016	Kerberos	Available
AWS RDS	2017	In-place Masking Mode	
Azure SQL	2019	Multi-tenant	Available
Azure Managed Instance	2022	Streams/Threads	Available

Platforms	Versions	Feature	Availability
Azure SQL Data Warehouse		Batch Update	Available
Google Cloud SQL Server		Drop Indexes	Available
		Disable Trigger	Available
		Disable Constraint	Available
		Identity Column Support	Available
		On-the-fly Masking Mode	
		Restart Ability	Available
		Truncate	Available
		Disable Trigger	Available
		Disable Constant	Available
		Profiling	
		Multi-tenant	Available
		Streams	Available

TLS/SSL setup

1. Add the database's certificate in the setup application using instructions in the **Adding a certificate** section, in the [TrustStore settings](#) article.
2. Restart the Compliance engine.
3. Create a MSSQL connector in Continuous Compliance with the relevant parameters. Upload a properties file for the connector with the following:

```

encrypt = true
trustServerCertificate = false / true <----- Depending upon whether to directly
accept the database certificate without checking certificate common-name
hostNameInCertificate = 10-110-229-143.qa-ad.delphix.com <----- This property
is only required in case trustServerCertificate is set to false

```

PostgreSQL connector

Introduction

PostgreSQL, often simply Postgres, is an object-relational database management system (ORDBMS) with an emphasis on extensibility and standards compliance. PostgreSQL is developed by the PostgreSQL Global Development Group, a diverse group of many companies and individual contributors. It is free and open-source, released under the terms of the PostgreSQL License, a permissive software license.

Support matrix

Platforms	Versions	Feature	Availability
Unix	9.2	TLS/SSL	Available
Linux	9.3	Password Vault	Available
Windows	9.4	Kerberos	Unavailable
AWS RDS	9.5	In-place Masking Mode	
AWS Aurora	9.6	Multi-tenant	Available
Azure Database for PostgreSQL	10	Streams/Threads	Available
Google Cloud SQL PostgreSQL	11	Batch Update	Available
	12	Drop Indexes	Unavailable
	13	Disable Trigger	Unavailable
	14	Disable Constraint	Unavailable
	Enterprise DB	Identity Column Support	Available
		On-the-fly Masking Mode	
		Restart Ability	Unavailable
		Truncate	Available
		Disable Trigger	Available
		Disable Constant	Available

Platforms	Versions	Feature	Availability
		Profiling	
		Multi-tenant	Available
		Streams	Unavailable

TLS/SSL setup

1. Add the database's certificate in the setup application using instructions in the **Adding a certificate** section, in the [TrustStore settings](#) article.
2. Restart the Compliance engine.
3. Create a PostgreSQL connector in Continuous Compliance with the relevant parameters. Upload a properties file for the connector with the following:

```
ssl=true
sslmode=verify-full
sslfactory=org.postgresql.ssl.DefaultJavaSSLFactory
```

Note: To use the verify-full setting (highest security), the PostgreSQL database certificate's Common Name (CN) field must match the hostname. To check the CN value in the certificate: `openssl x509 -in server.crt -text -noout`

MySQL / MariaDB connector

Introduction

MySQL is an open-source relational database management system (RDBMS). MySQL was owned and sponsored by a single for-profit firm, the Swedish company MySQL AB. MySQL is now owned by Oracle Corporation.

MariaDB is a community-developed fork of the MySQL relational database management system intended to remain free under the GNU GPL. Development is led by some of the original developers of MySQL, who forked it due to concerns over its acquisition by Oracle Corporation.

A MySQL Connector may be used to connect to either a MySQL or MariaDB database instance.

MySQL support matrix

Platforms	Versions	Feature	Availability
Unix	5.5	TLS/SSL	Available
Linux	5.6	Password Vault	Unavailable
Windows	5.7	Kerberos	Unavailable

Platforms	Versions	Feature	Availability
AWS RDS	8	In-place Masking Mode	
AWS Aurora		Multi-tenant	Available
Azure Database for MySQL		Streams/Threads	Available
Google Cloud SQL MySQL		Batch Update	Available
		Drop Indexes	Available
		Disable Trigger	Unavailable
		Disable Constraint	Unavailable
		Identity Column Support	Available
		On-the-fly Masking Mode	
		Restart Ability	Unavailable
		Truncate	Available
		Disable Trigger	Unavailable
		Disable Constant	Unavailable
		Profiling	
		Multi-tenant	Available
		Streams	Available

MariaDB support matrix

Platforms	Versions	Feature	Availability
Unix	10	TLS/SSL	Available
Linux		Password Vault	Unavailable

Platforms	Versions	Feature	Availability
Window		Kerberos	Unavailable
AWS RDS		In-place Masking Mode	
AWS Aurora		Multi-tenant	Available
Azure Database for MariaDB		Streams/Threads	Available
		Batch Update	Available
		Drop Indexes	Available
		Disable Trigger	Unavailable
		Disable Constraint	Unavailable
		Identity Column Support	Available
		On-the-fly Masking Mode	
		Restart Ability	Unavailable
		Truncate	Available
		Disable Trigger	Unavailable
		Disable Constant	Unavailable
		Profiling	
		Multi-tenant	Available
		Streams	Available

TLS/SSL setup

1. Add the database's certificate in the setup application using instructions in the **Adding a certificate** section, in the [TrustStore settings](#) article.
2. Restart the Compliance engine.

3. Create a MySQL/MariaDB connector in Continuous Compliance with the relevant parameters. Upload a properties file for the connector with the following:

```

useSSL=true
trustServerCertificate=false
clientCertificateKeyStoreUrl=file:/var/delphix/server/etc/.truststore
clientCertificateKeyStorePassword=changeit
clientCertificateKeyStoreType=JKS
trustCertificateKeyStoreUrl=file:/var/delphix/server/etc/.truststore
trustCertificateKeyStorePassword=changeit
trustCertificateKeyStoreType=JKS

```

SAP ASE (Sybase) connector

Introduction

SAP ASE (Adaptive Server Enterprise), originally known as Sybase SQL Server, and also commonly known as Sybase DB or Sybase ASE, is a relational model database server product for businesses developed by Sybase Corporation which became part of SAP AG.

Support matrix

Platforms	Versions	Feature	Availability
Unix	15.5	TLS/SSL	Available
Linux	15.7	Password Vault	Unavailable
Windows	16	Kerberos	Available
		In-place Masking Mode	
		Multi-tenant	Available
		Streams/Threads	Available
		Batch Update	Available
		Drop Indexes	Available
		Disable Trigger	Available
		Disable Constraint	Available
		Identity Column Support	Available

Platforms	Versions	Feature	Availability
		On-the-fly Masking Mode	
		Restart Ability	Available
		Truncate	Available
		Disable Trigger	Available
		Disable Constant	Available
		Profiling	
		Multi-tenant	Available
		Streams	Available

DB2 Z/OS and iSeries connectors

Introduction

DB2 for z/OS and iSeries are relational database management systems that run on IBM Z(mainframe) and IBM Power Systems.

Support matrix

i-Series	z/OS	Feature	Availability
7.1	11	TLS/SSL	Available
7.2	12	Password Vault	Unavailable
7.3		Kerberos	Unavailable
7.4		In-place Masking Mode	
		Multi-tenant	Available
		Streams/Threads	Available
		Batch Update	Available

i-Series	z/OS	Feature	Availability
		Drop Indexes	Unavailable
		Disable Trigger	Unavailable
		Disable Constraint	Unavailable
		Identity Column Support	Unavailable
		On-the-fly Masking Mode	
		Restart Ability	Unavailable
		Truncate	Available
		Disable Trigger	Unavailable
		Disable Constant	Unavailable
		Profiling	
		Multi-tenant	Available
		Streams	Available

Files connector

Introduction

Much of the time data will live outside of databases. The data can be stored in a variety of different formats including Fixed Width, Delimited, etc.

Support matrix

File Type/Format	Support Level
Fixed Width	Supported
Delimited	Supported
XML	Supported

File Type/Format	Support Level
JSON	Supported

Mainframe data set connector

Introduction

In addition to databases and files, the Masking Engine can process data stored in Mainframe data sets commonly found on the IBM z/OS operating system. For more information on data sets, see this [IBM knowledge center article](#).

Support matrix

The Masking Engine requires that data be encoded in EBCDIC rather than something like ASCII or UTF-8. EBCDIC is the encoding traditionally used on Mainframes.

On-The-Fly masking jobs

Continuous Compliance supports **On-The-Fly** (OTF) masking jobs where the data is read from a source location and written to a different target location. Only certain combinations of connector types are supported for OTF jobs.

OTF jobs with connectors of the same type are supported. For example, masking data from an Oracle source database to an Oracle target database is supported if both are using the built-in Oracle connector. OTF jobs using Extended Connectors are supported if both the source and target are using the same Extended Driver (the same uploaded JDBC driver). Additionally, OTF jobs with a relational database source and a delimited file target are supported. The following data sources are supported as source connectors for OTF jobs with delimited file targets.

- Oracle
- DB2
- MS SQL
- PostgreSQL
- MySQL / MariaDB
- SAP ASE (Sybase)
- Connectors created as [Extended Connectors](#).

For masking flat files (e.g. XML, delimited, etc) in an on-the-fly masking job, it is no longer required to copy or create empty files on the target. If the file name pattern does not match any file on the source, the execution will be reported as success, although no file is masked.

No other combinations of connector types are supported. For example, an Oracle source with a PostgreSQL target, or an MS SQL source with a fixed width file target, are unsupported.

Installation

This section covers the following articles:

- [Containerized installation](#)
- [Network connectivity requirements](#)
- [Prerequisites](#)
- [First time setup](#)
- [AWS EC2 installation](#)
- [Azure installation](#)
- [Google Cloud platform installation](#)
- [IBM Cloud platform installation](#)
- [Hyper-V installation](#)
- [OCI installation](#)
- [VMware installation](#)
- [Naming requirements](#)

Containerized installation

Kubernetes installation for containerized masking

This section describes how to utilize Delphix Kubernetes images to deploy a containerized version of our Continuous Compliance Engine. Continuous Compliance and Masking are used interchangeably throughout these documents.

With a few small exceptions, Containerized Masking provides the same functionality and user experience as when deployed as a Virtual Machine Masking Engine.

Obtaining the images

Containerized Masking utilizes 3 integrated containers to deliver essentially the same masking experience as our Virtual Machine Masking Engine. The containerized form allows for rapid spin up/tear down of ephemeral engines to handle automated workflow deployments. The 3 containers are delivered in a compressed archive (`.tar.gz`) for convenience.

Licensed versions of these bundles are available for download from the download.delphix.com site. In the folder for each version are 2 files. One file is HTML with instructions very similar to this page that can be downloaded to provide an offline copy of the installation instructions. The second file is the `masking_docker_images.tar.gz` bundle which contains the container images.

Docker is employed to build the container images which produces a set of [Open Source \(OCI\) images](#) for each container. The intention is to make the containers as vendor independent as possible.

Setup

Containerized Masking is intended to be run as a pod on Kubernetes. The pod consists of three containers:

1. `delphix-masking-app` - Serves the application UI and API, and executes masking jobs.
2. `delphix-masking-database` - Stores various application configuration.
3. `delphix-masking-proxy` - Serves as a reverse proxy handling HTTP and HTTPS traffic for the UI and API.

The API and UI are served from internal ports 8080 and 8443. When deploying the application, the kubernetes config must provide a Service which directs external HTTP traffic to port 8080 and HTTPS traffic to port 8443 as shown in the example `kubernetes-config.yaml` file.

The pod also requires a single volume per instance. This storage should be attached to both the app container and the database container.

- This volume should be attached to the `delphix-masking-database` container at location `/var/delphix/postgresql` with a subpath of `postgresql`.
- This volume should be attached to the `delphix-masking-app` container twice. Once at location `/var/delphix/masking/` with a subpath of `masking` and once at location `/var/delphix/postgresql` with a subpath of `postgresql`.

This volume should have at least 2GB of space for each container, though certain configurations may require significantly more space.

This storage volume should be created as a persistent volume. If it is not, masking job configurations will have to be recreated each time the pod is restarted. Also, certain diagnostic information captured in the logs will be lost when the pod is restarted unless the volume is persistent.

Because this volume is persistent, the pod should be deployed as a StatefulSet.

Network management

The proxy container has built-in configurations to act as a reverse proxy. It is recommended that the main `nginx.conf` file remains unmodified; instead, modify the individual component configuration files that get incorporated into the main `nginx.conf` file through include statements (such as `proxy.conf` for the reverse proxy-related configs and `ssl.conf` for HTTPS related configs).

To modify any nginx related files, such as config files or certificates and keys, an external volume should be bind mounted to the proxy container at `/etc/config`. During container startup, if the proxy container detects bind mounted files at the locations listed below, it will ignore the config files that are built into the proxy container's image and will instead use the mounted files.

HTTPS certificates

If the proxy container does not detect an external certificate in the expected location, it will generate and use a self-signed certificate.

The expected locations of each file are shown below:

File	Description
<code>/etc/config/nginx/nginx.conf</code>	main configs file
<code>/etc/config/nginx/proxy.conf</code>	reverse proxy configs
<code>/etc/config/nginx/ssl/ssl.conf</code>	ssl configs
<code>/etc/config/nginx/ssl/nginx.crt</code>	ssl certificate
<code>/etc/config/nginx/ssl/nginx.key</code>	ssl private key
<code>/etc/config/nginx/ssl/dhparam.pem</code>	DH parameters file

OWASP CSRFGuard

The [OWASP CSRFGuard](#) product has been employed as part of the protections that are built-in to the Masking product. The supplied NginX proxy container rewrites a packet's Host header with the contents of the X-Forwarded-Host header if it exists so that CSRFGuard will accept proxied packets.

This results in a requirement. If the Pod is placed behind a proxy device that re-writes the Host header, that proxy must add an X-Forwarded-Host header containing the original host value.

Sample configuration

The following configuration file shows an example of how Containerized Masking might be deployed. Details will vary based on the use case, environment, and product version.

```

apiVersion: v1
kind: Service
metadata:
  name: delphix-masking
spec:
  type: NodePort
  selector:
    app: masking
  ports:
    - name: http
      port: 8080
      nodePort: 30080
    - name: https
      port: 8443
      nodePort: 30443
---
apiVersion: apps/v1
kind: StatefulSet
metadata:
  name: delphix-masking
spec:
  selector:
    matchLabels:
      app: masking
  serviceName: delphix-masking
  template:
    metadata:
      labels:
        app: masking
    spec:
      securityContext:
        runAsUser: 65436 # masking user
        runAsGroup: 50 # staff group
        fsGroup: 50
        #
        # Some volume providers, such as hostProvider, do not support fsGroup.
        # If you are using such a volume provider, use an init container to
        # change the ownership of each volume to 65436:50 and the permissions
        # to 775.
        #
        runAsNonRoot: true
      containers:
        - image: delphix-masking-database:6.0.16.0-c1
          name: mds
          ports:
            - containerPort: 5432
              name: mds

```

```
volumeMounts:
  - name: masking-persistent-storage
    mountPath: /var/delphix/postgresql
    subPath: postgresql
- image: delphix-masking-app:6.0.16.0-c1
  name: app
  ports:
    - containerPort: 8284
      name: http
  volumeMounts:
    - name: masking-persistent-storage
      mountPath: /var/delphix/masking
      subPath: masking
    - name: masking-persistent-storage
      mountPath: /var/delphix/postgresql
      subPath: postgresql
  startupProbe:
    httpGet:
      scheme: HTTPS
      path: /masking/api/system-information
      port: 8443
    failureThreshold: 30
    periodSeconds: 10
    timeoutSeconds: 10
  livenessProbe:
    httpGet:
      scheme: HTTPS
      path: /masking/api/system-information
      port: 8443
    initialDelaySeconds: 300
    failureThreshold: 1
    periodSeconds: 10
    timeoutSeconds: 10
  readinessProbe:
    httpGet:
      scheme: HTTPS
      path: /masking/api/system-information
      port: 8443
    initialDelaySeconds: 30
    periodSeconds: 60
    timeoutSeconds: 10
- image: delphix-masking-proxy:6.0.16.0-c1
  name: proxy
  ports:
    - containerPort: 8080
      name: http
    - containerPort: 8443
      name: https
volumeClaimTemplates:
  - metadata:
      name: masking-persistent-storage
    spec:
```

```

accessModes:
  - ReadWriteOnce
resources:
  requests:
    storage: 4Gi

```

Deployment

Load the container images obtained from the download site into some Kubernetes container registry, then deploy the Masking Pod using a config file similar to the example provided above.

```
kubectl apply -f <path-to-config-file>
```

Debugging

In a support case, a Delphix Support engineer may ask for a support bundle containing diagnostic information. The preferred method of generating a support bundle is to use the API endpoints as shown in our document [API Call for Generating a Support Bundle](#).

Please see our [API Client Documentation](#) for more information regarding using the API Client.

Generating and retrieving a support bundle From the command-line

However, if the API endpoints are not functioning properly or there are difficulties accessing them, a support bundle can be gathered by running the following command-line commands from the Kubernetes layer of the node hosting the Pod: (Kubernetes admin permissions are required to perform these actions)

The exact name of the tarball created by this command can then be found using `kubectl exec`. For example:

```
$ kubectl exec -it <pod name> -c app -- /bin/bash /opt/delphix/masking/bin/
generate_container_support_bundle.sh
```

```
$ kubectl exec delphix-masking-0 -c app -- find /var/delphix/masking/ -name 'dlpx-
support-*'
/var/delphix/masking/dlpx-support-4b3e2af2-1d00-43f5-b45b-c84dba62648a-20211201-18-21-5
3.tar.gz
```

The tarball can then be copied out of the pod using `kubectl cp`. For example:

```
$ kubectl cp delphix-masking-0:/var/delphix/masking/dlpx-support-4b3e2af2-1d00-43f5-
b45b-c84dba62648a-20211201-18-21-53.tar.gz -c app dlpx-support-4b3e2af2-1d00-43f5-
b45b-c84dba62648a-20211201-18-21-53.tar.gz
```

The tarball can then be provided to the Delphix Support engineer by uploading it to upload.delphix.com and adding the associated case number in the matching field.

Network connectivity requirements

This topic covers the general network and connectivity requirements, including connection requirements, port allocation, and firewall and Intrusion Detection System (IDS) considerations.

i A security mechanism exists that does not allow the Masking engine to deploy behind a reverse proxy on the network.

General outbound connections from the virtual machine Delphix Continuous Compliance Engine

Protocol	Port Numbers	Use
TCP	25	Connection to a local SMTP server for sending email.
TCP/UDP	53	Connections to local DNS servers.
UDP	123	Connection to an NTP server.
UDP	162	Sending SNMP TRAP messages to an SNMP Manager.
TCP	443	HTTPS connections from the Delphix Engine to the Delphix Support upload server.
TCP/UDP	636	Secure connections to an LDAP server.
TCP/UDP	various	Connections to target environments such as databases (JDBC) and files (FTP, SFTP, NFS, or CIFS).

General inbound connections to the virtual machine Delphix Continuous Compliance Engine

Protocol	Port Numbers	Use
TCP	22	SSH connections to the Delphix Engine.
TCP	80	HTTP connections to the Delphix GUI (optional).
UDP	161	Messages from an SNMP Manager to the Delphix Engine.
TCP	443	HTTPS connections to the Delphix GUI.

General outbound connections from the containerized Delphix Continuous Compliance Engine

Containerized Masking is deployed as a Pod on a customer Kubernetes infrastructure rather than being a self-contained machine like the VM deployments. There is much underlying infrastructure (NTP, for example) that the VM deployment must manage, which is unnecessary for a containerized deployment. There are many features (again using time as one example) that a containerized deployment requires from the underlying infrastructure, but since they are no longer managed by the Pod itself, they no longer appear in the list of networking requirements.

Protocol	Port Numbers	Use
TCP	25	Connection to a local SMTP server for sending email.
TCP/UDP	53	Connections to local DNS servers.
TCP/UDP	various	Connections to target environments such as databases (JDBC) and files (FTP, SFTP, NFS, or CIFS).

General inbound connections to the containerized Delphix Continuous Compliance Engine

 The inbound ports shown in the table below are all internal. The kubernetes config defines a service that routes customer supplied external facing ports to the listed internal ports allowing the customer to choose any ports that work best for their infra. The example config maps external port 30080 to internal port 8080 and external port 30443 to internal port 8443, but that is left entirely to customer discretion.

Protocol	Port Numbers	Use
TCP	8080	HTTP connections to the Delphix GUI (optional).
TCP	8443	HTTPS connections to the Delphix GUI.

Firewalls and Intrusion Detection Systems (IDS)

Firewalls can add milliseconds to the latency between servers. Accordingly, for best performance, there should be no firewalls between the Delphix Masking Engine and the target environments. If the Delphix Masking Engine is separated from a target environment by a firewall, the firewall must be configured to permit network connections between the Delphix Masking Engine and the target environments for the application protocols (ports) listed above.

Intrusion detection systems (IDSs) should also be made permissive to the Delphix Masking Engine deployment. IDSs should be made aware of the anticipated high volumes of data transfer between the Delphix Masking Engine and target environments.

Prerequisites

VM-based Continuous Compliance Engines

This section will detail the hardware/software requirements needed to deploy the Delphix Engine with the Masking service. The Delphix Engine is a self-contained operating environment and application that is provided as a Virtual Appliance. Our Virtual Appliance is certified to run on a variety of platforms including VMware, AWS, and Azure.

The Delphix Engine should be placed on a server where it will not contend with other VMs for network, storage, or other computing resources. The Delphix Engine is a CPU and I/O-intensive application, and deploying it in an environment where it must share resources with other virtual machines can significantly reduce performance.

To use both Continuous Data (data virtualization) and Continuous Compliance (data masking), they must be deployed as separate Delphix engines. One Delphix engine is required per service, running both operations on one engine is not supported.

Client web browser

The Delphix Engine's graphical interface can be accessed from a variety of different web browsers. The Delphix Engine currently supports the following web browsers:

- Microsoft Edge 40.x or higher
- Mozilla Firefox 35.0 or higher
- Chrome 40 or higher

AWS EC2 platform

See [AWS EC2 Installation](#) for information about the virtual machine requirements for installation of a dedicated Delphix Masking Engine on Amazon's Elastic Cloud Compute (EC2) platform.

Azure platform

See [Azure Installation](#) for information about the virtual machine requirements for the installation of a dedicated Delphix Masking Engine on the Azure platform.

Google cloud platform

See [Google Cloud Platform Installation](#) for information about the virtual machine requirements for the installation of a dedicated Delphix Masking Engine on the GCP platform.

IBM Cloud

See [IBM Cloud Installation](#) for information about the virtual machine requirements for the installation of a dedicated Delphix Masking Engine on the IBM Cloud.

VMware platform

See [VMware installation](#) for information about the virtual machine requirements for the installation of a dedicated Delphix Masking Engine on the VMware Virtual platform.

Container (Kubernetes) based Continuous Compliance Engine

For Containerized Masking, the product is delivered as a set of containers that are deployed as a Pod in the customer's Kubernetes infrastructure. This Pod provides a very similar set of functionality as the Delphix Engine VM-based appliance.

Containerized Masking was developed to provide the ability to create ephemeral engines. I.e. small engines that can be spun up quickly for a specific need and then thrown away once that need is fulfilled.

The customer will need to provide said Kubernetes infrastructure whether on-prem infrastructure (such as MiniKube or MicroK8s) or cloud-based infrastructure. (such as AWS EKS)

Additionally, some functionality may require additional software to be installed on the Kubernetes node systems. For example, if the use of NFS-mounted filesystems is planned, each node would need the NFS client software to allow Kubernetes to perform the desired NFS mounts.

Differences from VMware-based Engines

The VMware-based Continuous Compliance engine was deployed conjoined with the Continuous Data product, including its Engine Setup App. The containerized version of Continuous Compliance is fully divorced from the Continuous Data product which means that some functionality that was provided or enabled by the Engine Setup App is not available. Some unavailable items are on the roadmap to be re-introduced to Containerized Continuous Compliance in future releases.

The ways in which the containerized service differs from its VMware-based compatriot are summarized in the following table.

Functionality	VMware based engine	Containerized engine
FTP support for file masking	Yes	No ¹
SSL / TLS for connectors	Yes ²	Yes ⁴
Local file masking via NFS	Yes ²	Yes ⁴
Local file masking via CIFS	Yes ²	Yes ⁴
LDAP	Yes ²	No ¹
Kerberos	Yes ²	No ¹
SSO / OAuth	Yes ²	No ³
Upgrade / upgrade validation	Yes ²	No
IBM's Custom Db2 Driver	Yes	No

-  1. Roadmap item, not currently supported.
 2. Via the Engine Setup app.
 3. Currently under evaluation.
 4. Via Kubernetes.

More information about each of the above can be found in the document sections that deal with their subject.

First time setup

This section walks you step by step on how to download and install the Delphix Engine software onto your infrastructure (VMware, AWS EC2, Azure, or GCP).

Setting up network access to the Delphix Engine

1. Power on the Delphix Engine and open the Console.
2. Wait for the Delphix Management Service and Delphix Boot Service to come online. This might take up to 10 minutes during the first boot. Wait for the large orange box to turn green.
3. Press any key to access the sysadmin console.
4. Enter **sysadmin** for the username and **sysadmin** for the password (when installing a new engine via AWS AMI, the initial sysadmin password is the AWS Instance ID).
5. You will be presented with a description of available network settings and instructions for editing.
6. Configure the hostname. Use the same hostname you entered during the server installation. If you are using DHCP, this step can be skipped.
7. Configure DNS. If you are using DHCP, this step can be skipped.
8. Configure either a static or DHCP address. The static IP address must be specified in CIDR notation (for example, 192.168.1.2/24).
9. Configure a default gateway. If you are using DHCP, this step can be skipped.
10. Commit your changes. Note that you can use the get command prior to committing to verify your desired configuration.
11. Check that the Delphix Engine can now be accessed through a Web browser by navigating to the displayed IP address, or hostname if using DNS.
12. Exit setup.

Setting up the Delphix Engine

Once you setup the network access for your Delphix Engine, enter the Delphix Engine URL in your browser for server setup. The Unified Setup wizard Welcome screen below will appear for you to begin your Delphix Engine setup.

Continuous Compliance Setup

Welcome

Choose engine type to setup:

Continuous Data

Continuous Compliance

This wizard will step you through the setup. During this process you will complete the following:

- Create your password for the default "sysadmin" user
- Set the system time
- Configure network and services
- Configure the storage pool
- Configure proxies, SMTP, and LDAP (these are optional)
- Register your software

After setup is complete, you will have two administrators defined:

- The system administrator, "**sysadmin**" with the password you defined. This will be the system administrator for the instance.
- The engine administrator, "**admin**" with the password you defined. This is typically a DBA who will administer all the data managed by the instance.
- The Continuous Compliance administrator, "**admin**" with the password you defined. This will be the Continuous Compliance administrator responsible for setting up users and other administrative actions in Continuous Compliance.

When setup is complete, log in as engine administrator to begin using your engine.

Back Next Submit

The Welcome page allows you to setup Masking-specific settings such as Masking admin user's email and password as well as Masking SMTP settings directly from the setup wizard. It will then redirect the customer to the corresponding login page based on the engine type selected.

When Masking is selected, the following will be added to the Welcome screen; "admin" with the password you defined. This will be the Masking administrator responsible for setting up users and other administrative actions in Masking.

There are limitations to this feature:

- Only Masking user settings (email and password) and SMTP settings are supported. Customers will need to use the API to setup LDAP.
- Once set, these settings can only be updated via the Masking API. There are no corresponding sections in the system dashboard.
- Engine Type cannot be modified once set in the Setup Wizard because it has other dependencies such as SSO.

 If the wrong password is entered, after 3 times the user will be locked out of the Masking service.

1. On the **Welcome** tab select **Masking** and then click **Next**.
2. In the Masking Password tab enter the current default (out-of-box) password for Masking. (Currently, the default is **Admin-12**)
3. Click **Validate** or **Next**. This causes the engine to validate the entered password with the masking service.
4. In the Administrators tab enter **System Administrator**, **Masking Administrator**, and **Engine Administrator** credentials. Then click **Next**.
5. Select an option for maintaining system time. Then click **Next**.
Note: The Masking engine only works with the UTC time zone. Time zone selection at the time of engine setup is not applicable for the Masking engine.
6. Configure your network interfaces and services and then select **Next**.
7. Delphix installs certificates signed by the Engines Certificate Authority. You can replace any certificate. Once you are ready click **Next**.
8. The Delphix Engine automatically discovers and displays storage devices. For each device, set the Usage Assignment to Data and set the Storage Profile to Striped. Then click **Next**.
9. Enter the **Masking SMTP** settings and then click **Next**.
10. The Authentication tab allows users to configure Virtualization LDAP settings. But Masking LDAP settings must be configured via the Masking API.
11. To enable SAML/SSO, set the Audience Restriction (SP entity ID, Partner's Entity ID) in the identity provider to be the Engine UUID. Select **Use SAML/SSO**. IdP metadata is an XML document which must be exported from the application created in your IdPCopy and pasted in the IdP Metadata field. Click **Next**.
12. If using Kerberos authentication select **Use Kerberos authentication** and complete all fields. Then enter **Next**.
13. If the Delphix Engine has access to the external Internet (either directly or through a web proxy), then you can auto-register the Delphix Engine. If external connectivity is not immediately available, you must perform manual registration. Copy the Delphix Engine registration code.
14. Click **Next**.
15. The final Summary tab will enable you to review your configuration. Click **Submit** to acknowledge the configuration.

Logging in to the Delphix Continuous Compliance Engine

1. Login to a web browser that points to `http://masking-engine.example.com/masking`.
2. Enter default username: `admin`.
3. Enter default user password: `Admin-12`.

AWS EC2 installation

This section covers the virtual machine requirements for installation of a dedicated Continuous Compliance Engine on Amazon's Elastic Cloud Compute (EC2) platform.

For best performance, the Continuous Compliance Engine and all database/file servers should be in the same AWS region.

The following topics are covered:

- Instance Types
- Network Configuration
- EBS Configuration
- General Storage Configuration
- Additional AWS Configuration Notes

Instance types

The Continuous Compliance Engine can run on a variety of different instances, including large memory instances (preferred) and high I/O instances. We recommend the following large memory and high I/O instances:

Requirements	Notes
Large Memory Instances: r5n.2xlarge r5n.4xlarge r5n.8xlarge r5n.16xlarge r5n.24xlarge r4.2xlarge r4.4xlarge r4.8xlarge r4.16xlarge	- Larger instance types provide more CPU, which can prevent resource shortfalls under high I/O throughput conditions. - Larger instances also provide more memory, which the Delphix Engine uses to cache database blocks. More memory will provide better read performance.
High I/O Instances (supported) i3.2xlarge i3.4xlarge i3.8xlarge	

 On the AWS EC2 platform, the Continuous Compliance Engine must have sufficient memory to operate when multiple masking jobs are running. Our recommendation is to provide 8 GB of memory for the Continuous Compliance Engine in addition to any memory that will be used by running jobs.

Network configuration

Requirements	Notes
Virtual Private Cloud	<ul style="list-style-type: none"> - You must deploy the Delphix Engine and all of the source and target environments in a VPC network to ensure that private IP addresses are static and do not change when you restart instances. - When adding environments to the Delphix Engine, you must use the host's VPC (static private) IP addresses.
Static Public IP	The EC2 Delphix instance must be launched with a static IP address; however, the default behavior for VPC instances is to launch with a dynamic public IP address – which can change whenever you restart the instance. If you're using a public IP address for your Delphix Engine, static IP addresses can only be achieved by using assigned AWS Elastic IP Addresses.
Security Group Configuration	The default security group will only open port 22 for SSH access. You must modify the security group to allow access to all of the networking ports used by the Delphix Engine and the various source and target engines.

Storage configurations

 You must always attach a minimum of 2 storage pools to the Delphix Engine; one for rpool and other for domain0 pool.

EBS configuration

Deploying Delphix on AWS EC2 requires EBS-provisioned IOPS volumes. Since EBS volumes are connected to EC2 instances via the network, other network activity on the instance can affect throughput to EBS volumes. EBS-optimized instances provide guaranteed throughput to EBS volumes and are required to provide consistent and predictable storage performance.

Requirements	Notes
<p>EBS Provisioned IOPS Volumes</p> <div style="background-color: #e6e6fa; padding: 5px; border: 1px solid #d1c4e9;"> <p> All attached storage devices must be EBS volumes.</p> </div>	<ul style="list-style-type: none"> - Delphix does not support the use of instance store volumes. - Use EBS volumes with provisioned IOPs in order to provide consistent and predictable performance. The number of provisioned IOPs depends on the estimated IO workload on the Delphix Engine. - Provisioned IOPs volumes must be configured with a volume size to provisioned IOPs per the EBS Volume Types guidelines. - I/O requests of up to 256 kilobytes (KB) are counted as a single I/O operation (IOP) for provisioned IOPs volumes. Each volume can be configured for up to 4,000 IOPs.

System disk

The minimum recommended storage size for the System Disk is 127 GB.

Metadata disk(s)

The minimum recommended storage size of the Metadata Volume is 50 GB.

General storage configuration

Requirements	Notes
<ul style="list-style-type: none"> - Allocate initial storage equal to the size of the physical source database storage. - Add storage when storage capacity approaches 30% free. 	<ul style="list-style-type: none"> - For high redo rates and/or high DB change rates, allocate an additional 10-20 %. - Add new storage by provisioning new volumes of the same size. - This enables the Delphix File System (DxFS) to make sure that its file systems are always consistent on disk without additional serialization. This also enables the Delphix Engine to achieve higher I/O rates by queueing more I/O operations to its storage.
EBS Volume Size and Count <ul style="list-style-type: none"> - Keep all EBS volumes the same size. Maximize Delphix Engine RAM for a larger system cache to service reads - Use at least 4 EBS volumes to maximize performance. 	This enables the Delphix File System (DxFS) to make sure that its file systems are always consistent on disk without additional serialization. This also enables the Delphix Engine to achieve higher I/O rates by queueing more I/O operations to its storage.

Additional AWS configuration notes

- Using storage other than EBS is not supported.
- Limits on the number of volumes are dictated by the EBS instance type, and is generally advised that over 40 can be expected to cause issue on Linux VMs. More information can be found in the [AWS Volume Limits](#) and [AWS Volume Constraints](#) articles. The maximum device limit imposed by AWS can be handled by the Delphix Engine.
- The use of the local SSDs attached to i2 instance types is not supported.
- Using fast storage for EBS volumes is supported and recommended, including (in order of decreasing speed):
 - Provisioned IOPS (io1) volumes (recommended).
 - Virtual Machine Requirements for AWS EC2 Platform
 - General Purpose SSD (gp2) volumes (supported)
 - Throughput Optimized HDD (st1) volumes (supported)
 - Cold HDD (sc1) volumes (not supported due to poor performance)
 - Magnetic (standard) volumes (supported, but use st1 instead where possible)

Installing AMI on AWS EC2

The following two methods can be used to install/deploy Continuous Compliance in AWS:

- Access Delphix provided AMI through the Delphix download site
- Subscribe to Continuous Compliance through the Amazon Marketplace

Using the Delphix download site to deploy masking

1. On the Delphix download site, click the AMI you would like to share and accept the Delphix License agreement. Alternatively, follow a link given by your Delphix solutions architect.

2. On the **Amazon Web Services Account** Details form presented:
 - a. Enter your **AWS Account Identifier**, which can be found here: <https://console.aws.amazon.com/billing/home?#/account>. If you want to use the GovCloud AWS Region, be sure to enter the ID for the AWS Account which has GovCloud enabled.
 - b. Select which **AWS Region** you would like the AMI to be shared in. If you would like the AMI shared in a different region, contact your Delphix account representative to make the proper arrangements.
3. Click **Share**. The Delphix Engine will appear in your list of AMIs in AWS momentarily.
4. Reference the Installation and Configuration Requirements for AWS/EC2 when deploying the AMI.
5. Once you have launched your Continuous Compliance EC2 instance and it is accessible via a web browser (port 80), proceed to [First time setup](#) to configure the system.

Subscribing to Continuous Compliance through the Amazon Marketplace

1. Sign in to the AWS Console.
2. Navigate to AWS Marketplace.
3. Typing Delphix in the search bar will find several Delphix Product offerings. Select **Continuous Compliance for AWS (3TB)**.
4. Click **Continue to Subscribe**.
5. Click **Accept Terms**.
6. Wait for the subscription to be confirmed, then click **Continue to Configuration**.
7. Select or verify the correct **Region** for launch/deployment.
8. Then click **Continue to Launch**.
9. Select either to **Launch from Website** or **Launch through EC2**.
10. For either option you will need to enter the following:
 - a. VPC in which to launch the instance.
 - b. Subnet on which the instance will reside.
 - c. Instance Type (Recommended: r4.2xlarge).
 - d. Security Group (Minimal access required: 22, 80, or 443)
11. Once the Delphix EC2 instance is launched proceed to [Setting up the Delphix Engine](#) to configure the system.

Azure installation

This topic covers the virtual machine requirements, including memory and data storage, for deploying the Delphix Engine on the Azure public cloud and Government Cloud.

Instance types

The Delphix Engine can run on a variety of different Azure instances. We recommend the following instances:

Requirements	Notes
Memory-Optimizes	
DS14v2 E8S_v3 E16S_v3 E32S_v3	16 CPUs, 112GB, 32 devices 8 CPUs, 112GB, 16 devices 16 CPUs, 244GB, 32 devices 32 CPUs, 448GB, 64 devices Network bandwidth and IOPS limits are specific to each instance type: - See DSv2 specifications for more details. - See GS specifications for more details.
General Purpose	
D16s_v3 D32_v3	Network bandwidth and IOPS limits are specific to each instance type: - See DSv2 specifications for more details. - See https://docs.microsoft.com/en-us/azure/virtual-machines/dv3-dsv3-series for more details.



On the Azure platform, recommendation is to provide 8 GB of memory for the Delphix Masking Engine in addition to any memory that will be used by running jobs.

Network configuration

Requirements	Notes
Azure Virtual Network (VNet)	The Delphix Engine and all the source and target environments must be accessible within the same virtual network.
Network Security Group (NSG)	You must modify the security group to allow access to all of the networking ports used by the Delphix Engine and the various source and target platforms.

See [Network connectivity requirements](#) for information about specific port configurations.

Storage configuration

 You must always attach a minimum of 2 storage pools to the Delphix Engine; one for rpool and other for domain0 pool.

We recommend using a total of four disks to run your Delphix Engine. One disk is used for the Delphix File System (DxFS) to ensure that its file systems are always consistent on disk without additional serialization. The other three disks will be used for data storage. This also enables the Delphix Engine to achieve higher I/O rates by queueing more I/O operations to its storage.

Requirements	Notes
Azure Premium Storage	<ul style="list-style-type: none"> - Premium storage utilizes solid-state drives (SSDs) - Devices up to 4096GB are supported - Maximum of 256TB is supported - I/O requests of up to 256 kilobytes (KB) are counted as a single I/O operation (IOP) for provisioned IOPS volumes - IOPS vary based on storage size with a maximum of 7,500 IOPS
System Disk	The minimum recommended storage size for the System Disk is 127 GB.
Metadata Disk(s)	The minimum recommended storage size of the Metadata Volume is 50 GB.

Extensions

Extensions are not currently supported.

Installing VHD on AZURE

Use the following steps to install your VHD:

1. On the [Microsoft Azure Marketplace](#), search for Delphix. Click **GET IT NOW**.
2. Reference the Installation and Configuration Requirements for the Delphix Engine in Azure when deploying the VHD.
3. Jump to [Setting up the Delphix Engine](#) section to learn how to activate the masking service now that you have the software installed.

Google Cloud Platform installation

This section covers the virtual machine requirements for the installation of a dedicated Continuous Compliance Engine on Google Cloud Platform (GCP).

Machine types

The following is a list of instance types that are supported to deploy Delphix on GCP. Delphix periodically certifies new instance types, which will be added to the list here.

Requirements	Notes
n2-standard-(16, 32, 64)	Larger instance types provide more CPU, which can prevent resource shortfalls under high I/O throughput conditions.
n2-highmem-(8, 16, 32, 64)	Larger instances also provide more memory, which the Delphix Engine uses to cache database blocks. More memory will provide better read performance.

Network configuration

Requirements	Notes
Virtual Private Cloud	You must deploy the Delphix Engine and database/file servers in a VPC network to ensure that private IP addresses are static and do not change when you restart instances. When adding connectors to the Masking Engine, you must use the host's VPC (static private) IP addresses.
Static Public IP	The GCP Delphix instance must be launched with a static IP address; however, the default behavior for VPC instances is to launch with a dynamic public IP address – which can change whenever you restart the instance.
Security Group Configuration	The default security group will only open port 22 for SSH access. You must modify the security group to allow access to all of the networking ports used by the Delphix Engine and the various source and target engines.
Premium Networking	It is recommended to use GCP Premium Tier Networking.

Storage configuration

 You must always attach a minimum of 2 storage pools to the Delphix Engine; one for rpool and other for domain0 pool.

System disk

The minimum recommended storage size for the System Disk is 127 GB.

Metadata disk(s)

The minimum recommended storage size of the Metadata Volume is 50 GB.

Additional GCP configuration notes

- Delphix supports both Zonal and Regional SSD persistent disks.

Installing on Google Cloud Platform

This section covers the requirements, including memory and data storage, for deploying the Delphix Engine on the Google Cloud Platform (GCP).

Prerequisites to deploying in GCP

- A license is required to use the Delphix software. If you are a new customer contact Delphix to get started.

Deploying a Delphix Engine in GCP

1. Log into Google Cloud Marketplace with your account.
2. Search for **Delphix**.
3. Click **Launch on Compute Engine**.
 - Machine Type: See the table below for supported configurations.
 - Boot disk type: SSD Persistent Disk
 - Boot disk size in GB: 127
 - Networking interfaces: Configure as appropriate for your environment
 - IP forwarding: Configure as appropriate for your environment
4. Click on **Deploy**.
5. Once deployed, go to [Setting up the Delphix Engine](#) section to learn how to activate the masking service now that you have the software installed.

IBM Cloud Platform installation

This topic covers the virtual machine requirements, including memory and data storage, for the deployment of the Delphix Engine on IBM Cloud.

Supported profiles

The following is a list of profiles that are supported to deploy Delphix on IBM Cloud.

Requirements	Notes
mx2-8x64 mx2-16x128 mx2-32x256 mx2-48x384	<ul style="list-style-type: none"> - The Delphix Engine most closely resembles a storage appliance and performs best when provisioned using a storage-optimized profile - Larger profiles provide more CPU, which can prevent resource shortfalls under high I/O throughput conditions. - Larger profiles also provide more memory, which the Delphix Engine uses to cache database blocks. More memory will provide better read performance.

Network configuration

Requirements	Notes
Virtual Server Instances	<ul style="list-style-type: none"> - You must deploy the Delphix Engine and all of the source and target environments in the same VPC network. - When adding environments to the Delphix Engine, you must use the host's VPC IP addresses.
Security Configuration	<ul style="list-style-type: none"> - The default security group will only open port 22 for SSH access. You must modify the security group to allow access to all of the networking ports used by the Delphix Engine and the various source and target engines. - See Network Performance Configuration Options for information about network performance tuning. - See General Network and Connectivity Requirements for information about specific port configurations. - Reference: IBM Cloud Security and Compliance documentation

Storage configuration



You must always attach a minimum of 2 storage pools to the Delphix Engine; one for rpool and other for domain0 pool.

Requirements	Notes
<ul style="list-style-type: none"> - Allocate initial storage equal to the size of the physical source database storage. - Add storage when storage capacity approaches 30% free. 	<ul style="list-style-type: none"> - For high redo rates and/or high DB change rates, allocate an additional 10-20 %. - Add new storage by provisioning new volumes of the same size. This enables the Delphix File System (DxFS) to make sure that its file systems are always consistent on disk without additional serialization. This also enables the Delphix Engine to achieve higher I/O rates by queueing more I/O operations to its storage. - A Delphix Engine requires a minimum of three (3) equally sized Block Volumes, in addition to the Boot volume which was automatically created while creating the virtual server instance. - IBM Block Storage Documentation

Additional IBM configuration notes

- Resize/expansion of a storage volume
- Expandable volume is a beta feature that is available for evaluation and testing purposes. This feature is available in the US South, US East, London, and France regions. Contact your IBM Sales representative if you are interested in getting access to [Expanding Block Storage](#).
- After performing an “online” resize/expansion of a storage volume using IBM Cloud tools, then use the Delphix sysadmin interface to “Expand” the storage device; otherwise, the newly allocated storage space, from the resize/expansion, will not be used.
- Resize/expansion of a storage volume using IBM Cloud is not supported while the Delphix Engine is in a stopped state.
- Removing a storage volume
- It should be done while the machine is running.
- First, use the Delphix sysadmin CLI interface to “Unconfigure” the storage device, then remove it from IBM Cloud.

Procedure for deploying in the IBM Cloud

Prerequisites to Deploying in IBM Cloud

1. You require a license to use Delphix software. If you are a new customer, contact [Delphix](#) to get started.
2. Review [IBM’s cloud documentation](#) for IBM Cloud-specific information.

Deploying Delphix in the IBM Cloud

There are two methods for deploying a Delphix Engine in the IBM Cloud using the Software Catalog or Manually Uploading the Delphix Image.

Deploying from the IBM Software Catalog

1. Navigate to the [IBM Software Catalog](#) and search for Delphix.
2. Select the Delphix Data Masking Tile for the Masking product.
3. Scroll down to the Deployment Values section and input the specifics for your environment.

Required Parameters	Description
hostname	The name of the VSI you will use to deploy Delphix.

Required Parameters	Description
profile	Compute profile to be used for deploying Delphix (see recommended profiles).
ssh_key	Your public SSH key to be used when provisioning the VSI.
subnet_id	The id of the subnet where the VSI will be provisioned.
volumecount	Number of block storage volumes.
volumeprofile	Block storage profile to use (recommended is >= 10 IOPS/GB)
volumesize	Block storage volume size.
vpcname	The name of your VPC where the VSI is provisioned.
zone	VPC zone to provision your environment.

Manually downloading and deploying the Delphix Image

Downloading the Delphix Image

 Contact your account manager to request access to the IBM variant of the Delphix product.

1. Follow the link given to you by your Delphix solutions architect. Download the Delphix_Verson...._Standard_IBM.qcow2 file and the SHA256SUMS file.
2. Once both files have finished downloading and assuming both files were downloaded to the same directory, you can run the following command to verify the download:

```
$ grep -i IBM.qcow2 ./SHA256SUMS | sed -E 's,Appliance_Images/(Controlled_Availability)?, ,g' | sha256sum --check
```

Uploading the Delphix Engine image as an object

1. Authenticate with the IBM Cloud and navigate to the <https://cloud.ibm.com/login/>.
2. Use the navigation menu to reach the **Resource List** page. The Resource List page can be navigated from the Dashboard by clicking on **Storage** within the **Resource summary** pane.
3. Expand Storage from the menu and select the appropriate resource group. You should have [created a resource group](#) depending on your organization's strategy for managing IBM resources.
4. [Create a storage bucket](#) or select an existing bucket.
5. Click the blue **Upload** button and select **Files**.
6. A pop-up menu appears to select the transfer type. **Aspera High-Speed Transfer** is required for large files. For this, you will need to install the plugin. It will automatically navigate you through the steps to install the plugin.
7. In the **Upload Files (objects)** window, click on the **Select Files (objects)** button and choose the IBM specific **QCOW2** file that was previously downloaded.
8. Click the **Upload** button.

Creating a custom image

1. Authenticate with IBM Cloud and navigate to the [Dashboard](#).
2. Use the navigation menu to reach the **Custom images** page for VPC within the VPC infrastructure (IBM Cloud pull-down menu, upper left, VPC Infrastructure > Custom images).
3. Click the blue **Create** button.
4. In the **Import custom image** page, specify a unique name for the image.
5. From the **Resource Group** drop-down, select your organization's resource group.
6. Optional: In the **Tags** section, provide appropriate [tags](#) to organize your resources.
7. Select the appropriate **Region**.
8. Select the **Cloud Object Storage** bucket containing the uploaded image by selecting the appropriate **Cloud Object Storage instances > Location > Bucket** from the drop-down menus. The downloaded QCOW2 image should appear in the pane below the three drop-down menus.
9. Within the **Operating System** section, click on the **Ubuntu Linux** tile and select **ubuntu-18-04-amd64** from the drop-down menu.
10. Once all the parameters are entered, in the right pane click on the blue button to **Import custom image**.

Launching the Delphix Engine

1. Authenticate with IBM Cloud and navigate to the [Dashboard](#).
2. Use the navigation menu to reach the **Virtual Server Instances** page within the VPC Infrastructure (IBM Cloud pull-down menu, upper left, VPC Infrastructure > Virtual Server Instances). **Note:** To maximize performance, deploy the Delphix Engine instance in the same VPC/subnet in which you will create your virtual databases (VDBs).
3. Click the blue **Create** button.
4. In the **New Virtual Server for VPC** page, specify a unique name for the VM.
5. From the **Virtual Private Cloud** drop-down, select your organization's VPC.
6. From the **Resource Group** drop-down, select your organization's resource group.
7. Optional: In the **Tags** section, provide appropriate [tags](#) to organize your resources.
8. Select the Location of your IBM Cloud resources.
9. In the **Operating System** section, click on the **Select Custom Image** link within the **Custom Image** block.
10. In the pop-menu, select the IBM-specific image you previously uploaded.
11. Within the **Profile** section, click on **View all profiles**. Select one of the supported instance types and click Save.
12. You can skip the **User Data** section.
13. You can also skip the **Boot Volume** section since it would already have the default values.
14. You can create block storage volumes later, so skip that for now. It will be discussed in the next section.
15. Continue on to the **Network Interfaces** section. If you already have a subnet configured in your zone and VPC, then this section will already have a default network interface. Otherwise, you need to create a subnet with the appropriate security groups. This part is critical, if the network isn't specified correctly, you are likely to run into firewall issues; please consult your IT or DevOps teams. Configure Network Security Groups (NSGs) for your subnet as required; again, please consult your IT or DevOps teams.
16. Click the **Create virtual server instance** button on the right panel. This will take a couple of minutes.

Creating block storage volumes

1. Authenticate with IBM Cloud and navigate to the [Dashboard](#).
2. Use the navigation menu to reach the **Block Storage Volumes** within VPC Infrastructure (IBM Cloud pull-down menu > VPC Infrastructure > Block Storage Volumes).
3. Click the blue **Create** button.
4. In the **Block Storage Volume for VPC** modal window, specify a unique name for this Block Volume. It can be helpful if this name is descriptive or identifies the VM it is intended to be attached to and ends in a sequence number.
5. From the **Resource Group** drop-down, select your organization's resource group.
6. Optional: In the **Tags** section, provide appropriate [tags](#) to organize your resources.

7. Select the Location of your IBM Cloud resources.
8. Enter the required IOPS. The recommended supported IOPS is 10/GB.
9. Enter the storage size in GB. Set the size of the volume to be sufficiently large, with room for growth, to support the databases that will be virtualized, or masked, by this Delphix Engine.
10. For **Encryption**, you can let it be the default, e.g. **Provider Managed**.
11. Click the blue **Create** Volume button. A Delphix Engine requires a minimum of three (3) equally sized Block Volumes, in addition to the Boot volume which was automatically created while creating the virtual server instance. Repeat Steps 3-11 as many times as necessary.

Attaching block storage volumes

1. Authenticate with IBM Cloud and navigate to the [Dashboard](#).
2. Use the navigation menu to reach the **Block Storage Volumes** within VPC Infrastructure (IBM Cloud pull-down menu > VPC Infrastructure > Block Storage Volumes).
3. From the list of pre-existing Block Volumes, identify the volumes you wish to attach to a Delphix Engine and wait until the volume's state becomes Available.
4. Note that the volumes you wish to attach have **Attachment Type** set as a **hyphen**.
5. The right side of the volume row shows an Expandable menu. Click on it and select **Attach to Instance**.
6. In the **Attach Virtual Server Instance** modal window, select your virtual server instance (Delphix Engine) from the drop-down menu.
7. Click on the blue **Attach Volume** button.
8. Repeat Steps 3-7 until all associated Block Volume resources have been attached to the Delphix Engine instance.

Configuring the Delphix Engine

1. Connect to your running Delphix Engine instance with a web browser. Use the IP address or DNS name noted in the Instance Description. Upon successful connection, the browser will display a login prompt to enter the Delphix Setup Page.
2. Refer to the standard product deployment instructions to complete your Delphix deployment.

Next Steps

Congratulations! You have successfully deployed a Delphix Engine in IBM Cloud.

Hyper-V installation

The Delphix Engine is a virtual appliance that runs in a hypervisor. In this section, you'll find requirements to run Delphix on Hyper-V including supported versions and instance configurations as well as recommended configuration parameters for optimal performance.

Contact your Delphix representative to request this capability. Delphix will assist you to review that all Hyper-V requirements are met to successfully run a Delphix Engine with the most appropriate configuration for your Use Cases.

If the Delphix Engine competes with other virtual machines on the same host for resources it will result in increased latency for all operations. As such, it is crucial that your Hyper-V host is not over-subscribed, as this eliminates the possibility of a lack of resources for the Delphix Engine. This includes allowing a percentage of CPU resources for the hypervisor itself as it can de-schedule an entire VM if the hypervisor is needed for managing IO or compute resources.

Supported versions

- Hyper-V Version: 10.0 and later
- Gen 1 only is supported

Virtual CPUs

Requirements	Notes
8 vCPUs	<ul style="list-style-type: none"> - CPU resource shortfalls can occur both on an over-committed host as well as competition for host resources during high IO utilization. - CPU reservations are strongly recommended for the Delphix VM so that Delphix is guaranteed the full complement of vCPUs even when resources are overcommitted. - It is suggested to use a single core per socket unless there are specific requirements for other VMs on the same Hyper-V host.
Never allocate all available physical CPUs to virtual machines	<ul style="list-style-type: none"> - CPU for the Hyper-V Server to perform hypervisor activities must be set aside before assigning vCPUs to Delphix and other VMs. - We recommend that a minimum of 8-10% of the CPUs available are reserved for hypervisor operation. (e.g. 12 vCPUs on a 128 vCore system).

Memory

Requirements	Notes
<p>128 or higher GB vRAM (recommended) 64GB vRAM (minimum)</p>	<ul style="list-style-type: none"> - The masking service on the Delphix Engine uses its memory to process database and file blocks. - Memory reservations are required for the Delphix VM. The performance of the Delphix Engine will be significantly impacted by the over-commitment of memory resources in the Hyper-V Server. - Reservations ensure that the Delphix Engine will not be forced to swap pages during times of memory pressure on the host. A swapped page will require orders of magnitude more time to be brought back to physical memory from the Hyper-V swap device.
<p>Memory for the Hyper-V Server to perform hypervisor activities must be set aside before assigning memory to Delphix and other VMs.</p>	<p>Failure to ensure sufficient memory for the host can result in a hard memory state for all VMs on the host which will result in a block for memory allocations.</p>

Network

Requirements	Notes
<p>Virtual ethernet adapter requirements.</p>	<ul style="list-style-type: none"> - SR-IOV recommended for all virtual ethernet adapters that will be used for Delphix data IO. - Jumbo frames recommended. - A 10GbE NIC in the Hyper-V Server is recommended.
<p>If the network load in the Hyper-V Server hosting the Delphix engine VM is high, dedicate one or more physical NICs to the Delphix Engine.</p>	<ul style="list-style-type: none"> - Adding NICs only works if VMs are discovered using different interfaces.

SCSI Controller

Requirements	Notes
LSI Logic Parallel	<p>- Per Hyper-V Storage I/O Performance Tuning Guidelines, it is recommended that you attach multiple disks to a single virtual SCSI controller and create additional controllers only as they are required to scale the number of disks connected to the virtual machine. For example, a VM with 3 virtual disks should distribute the disks across the single SCSI controller as follows:</p> <ul style="list-style-type: none"> - IDE Controller 1 - Boot Drive - SCSI Controllers - Disk 1, Disk 2, Disk 3 <div style="background-color: #e6e6ff; padding: 5px; margin-top: 10px;"> <p> For load purposes, we generally focus on the DB storage and ignore the controller placement of the system disk.</p> </div>

Storage configuration

-  You must always attach a minimum of 2 storage pools to the Delphix Engine; one for rpool and other for domain0 pool.

Requirements	Notes
Storage used for Delphix must be provisioned from storage that provides data protection.	<p>For example, using RAID levels with data protection features, or equivalent technology.</p> <p>The Delphix Engine does not protect against data loss originating at the hypervisor or SAN layers.</p>

Delphix storage operations

There are three types of data that Delphix stores on disk, which are:

1. Delphix VM Configuration Storage: stores data related to the configuration of the Delphix VM. VM Configuration Storage includes the Hyper-V configuration data as well as log files.
2. Delphix Engine System Disk Storage: stores data related to the Delphix Engine system data, such as the Delphix .ova settings.

Delphix VM configuration storage

The Delphix VM configuration must be stored on an NTFS volume(s).

Requirements	Notes
The volumes should have enough available space to hold all Hyper-V configuration and log files associated with the Delphix Engine.	If a memory reservation is not enabled for the Delphix Engine (in violation of memory requirements stated above), then space for a paging area equal to the Delphix Engine's VM memory must be added to the volumes containing the Delphix VM configuration data.

Delphix Engine system disk storage

Requirements	Notes
The Delphix Engine disks must be stored on NTFS volume(s).	The volume for the Delphix Engine System Disk Storage is often created on the same volume as the Delphix VM definition. In that case, the volume must have sufficient space to hold the Delphix VM Configuration, the virtual disk for the system disk, and a paging area if a memory reservation was not enabled for the Delphix Engine.
The Delphix .vhdx file is configured for a 128GB system drive.	The volume where the .vhdx is deployed should, therefore, have at least 128GB of free space prior to deploying the .vhdx.

Metadata disk(s)

In addition to making sure the latest Hyper-V patches have been applied, check with your hardware vendor for updates specific to your hardware configuration. VHDXs (virtual machine disks).

Requirements	Notes
A minimum of 4 VHDXs should be allocated for database storage.	Allocating a minimum of 4 VHDXs for database storage enables the Delphix File System (DxFS) to make sure that its file systems are always consistent on disk without additional serialization. This also enables the Delphix Engine to achieve higher I/O rates by queueing more I/O operations to its storage.
If using VHDXs: <ul style="list-style-type: none"> - Each VHDX should be the only VHDX on its NTFS volume - The VHDX volumes should be assigned to dedicated physical LUNs on redundant storage. - The VHDXs should be created as the Fixed Size type. 	Provisioning VHDXs from isolated volumes on dedicated physical LUNs: <ul style="list-style-type: none"> - Reduces contention for the underlying physical LUNs - Eliminates contention for locks on the volumes from other VMs and/or the Hyper-V Server Console
The quantity and size of VHDXs or RDMS assigned must be identical across all 4 controllers.	If the underlying storage array allocates physical LUNs by carving them from RAID groups, the LUNs should be allocated from different RAID groups. This eliminates contention for the underlying disks in the RAID groups as the Delphix engine distributes IO across its storage devices.
The physical LUNs used for NTFS volumes and RDMS should be of the same type in terms of performance characteristics such as latency, RPMs, and RAID level.	The total number of disk drives that comprise the set of physical LUNs should be capable of providing the desired aggregate I/O throughput (MB/sec) and IOPS (Input/Output Operations per Second) for all virtual databases that will be hosted by the Delphix Engine.

Requirements	Notes
The physical LUNs used for NTFS volumes can be thin-provisioned in the storage array.	If the storage array allocates physical LUNs from storage pools comprising dozens of disk drives, the LUNs should be distributed evenly across the available pools.

Installing Hyper-V

1. Download the image from Delphix's Download site and copy it to your VM directory.
2. Start the Hyper-V Manager and specify **Name and Location** and then select **Next**.
3. Specify the **Generation**, configure memory, and then select **Next**. Memory: 64 GB (minimum), 128 GB (recommended)
4. Set up Networking by selecting **vNIC** then select **Next**.
5. Attach the downloaded image as a boot disk. Create a unique boot disk for each image.

Note:
Boot disks cannot be shared.

 - Use an existing virtual hard disk.
 - Browse to the location of VM.
 - Select the Image.
6. Select **Finish**, the VM will appear in the inventory.
7. Customize the VM by selecting Settings:
 - Delphix recommends having the IDE be the first device to boot from (under BIOS setting).
 - Adjust the number of CPU (min 8)
 - Add Hard Drive. Use VHDX formatted disks. Recommend Fixed Size.

Note:
Differencing Disk Types are not supported.

 - 128 GB Disk Storage
8. Repeat step 7 as necessary.
9. Connect to the console and start the VM.
10. Once the installation is complete go to [Setting up the Delphix Engine](#) section to learn how to activate the masking service now that you have the software installed.

OCI installation

This topic covers the virtual machine requirements for deploying the Continuous Compliance Engine on Oracle Cloud Infrastructure (OCI).

Supported databases

Oracle databases up to version 19c are supported. Please reference the [Oracle Support Matrix](#) for the detailed list.

Compute image types

Delphix distributes product images, for OCI, using the QCOW2 image type. Compute Images must be imported into OCI using the Paravirtualized launch mode; currently, images using the Emulated launch mode are not supported.

Supported shapes

The following is a list of shapes that are supported to deploy Delphix on OCI.

Requirements	Notes
Large Memory Instances (perferred) VM.Standard2.8 VM_Standard2.16 VM_Standard2.24	The Delphix Engine most closely resembles a storage appliance and performs best when provisioned using a storage-optimized shape. Larger shapes provide more CPU, which can prevent resource shortfalls under high I/O throughput conditions. Larger shapes also provide more memory, which the Delphix Engine uses to cache database blocks. More memory will provide better read performance.

Network configuration

Requirements	Notes
Virtual Cloud Network (VCN)	You must deploy the Delphix Engine and all of the source and target environments in a VCN to ensure that private IP addresses are static and do not change when you restart instances. By default, OCI subnets are considered public. When defining a subnet, we encourage configuring it as private. Unless required by your environment, your VCN should not include a Public Subnet. When adding environments to the Delphix Engine, you must use the host's VCN (static private) IP addresses.
Static Private IP	The Delphix instance should be launched with a static private IP address. For security reasons, it is encouraged to avoid configuring your engine with a Public IP address; but, in some cases, it may be ok to use a dynamic Public IP address in addition to a static Private IP address if your environment requires such access.

Requirements	Notes
Security Rules Configuration	<p>OCI allows two firewall features: Network Security Groups (NSGs) and Security Lists. Oracle recommends the use of NSGs over Security Lists because “NSGs let you separate the VCN's subnet architecture from your application security requirements.”</p> <p>However, a VCN will use a Security List to define default rules. By default, the security list will only open port 22 for SSH access. You must modify the security list, or create NSGs, to allow access to all of the networking ports used by the Delphix Engine and the various source and target engines.</p> <p>This dual implementation of firewall, or security, rules may be different from other clouds. Please see OCI documentation for best practices.</p> <p>See Network Connectivity Requirements for information about specific port configurations.</p>

Storage configuration

 You must always attach a minimum of 2 storage pools to the Delphix Engine; one for rpool and other for domain0 pool.

Requirements	Notes
<p>Allocate initial storage equal to the size of the physical source database storage.</p> <p>Attach a minimum of four (4), equally sized, storage devices to the Delphix Engine.</p> <p>Add storage when storage capacity approaches 30% free.</p> <p>Must use Block Volume for data storage.</p> <p>Block Volumes must be attached using Paravirtualized mode.</p>	<p>Currently supported Instance Types, or Shapes, only support Block Volumes; File Storage is not supported.</p> <p>Paravirtualized block devices are required; currently, iSCSI devices are not supported.</p> <p>Elastic Performance Configuration Options (aka Volume Performance Policy): use Higher Performance.</p> <p>For high redo rates and/or high DB change rates, allocate an additional 10-20 %.</p> <p>Add new storage by provisioning new volumes of the same size. This enables the Delphix Engine to achieve higher I/O rates by distributing load among devices and queueing more I/O operations to its storage.</p>

Additional OCI configuration notes

- When running low on storage space, Delphix recommends adding additional equivalently sized block storage volumes, or devices, instead of resizing existing volumes.
- If you must expand existing storage volumes, then this must be done using the “online” resizing strategy specified in OCI documentation; “offline” storage resizing is not supported and may lead to unexpected downtime. If an existing storage volume is expanded, then use the Setup, or sysadmin, interface to expand each storage “device,” or volume. The additional storage, as a result of a resize, will not be available for use until the storage devices are explicitly instructed to make use of the additional space.

- If expanding storage volumes, it is recommended that all volumes are expanded to the same size. When storage volumes, or devices, are the same size the Delphix product is able to balance I/O distribution among the disks for optimal performance.
- Hot removal of storage volumes is not supported.

Installing OCI

Download and verify the Delphix Engine image

1. Contact your account manager to request access to the OCI variant of the Delphix product.
2. Follow the link given by your Delphix solutions architect. Download the `Delphix_6.x.x.x_...Standard_OCI.qcow2` file and the `SHA256SUMS` file.
3. Once both files have finished downloading and assuming both files were downloaded to the same directory, you can run the following command to verify the download:

```
$ grep -i OCI.qcow2 ./SHA256SUMS | sed -E 's,Appliance_Images/(Controlled_Availability/)?, ,g' | sha256sum --check
```

Upload the Delphix Engine image as an object

1. Authenticate with OCI and navigate to the [Infrastructure Console](#).
2. Use the navigation menu to reach the **Object Storage Buckets, Core Infrastructure**, page (Hamburger Menu > Object Storage > Object Storage).
3. Remember to set your **List Scope Compartment**. This will depend on your organization's strategy for managing OCI resources.
4. [Create a storage bucket](#) or select an existing bucket.
5. Click the blue **Upload** button.
6. In the **Upload Objects** modal window, specify an optional prefix and choose the OCI specific QCOW2 file that was previously downloaded.
7. Click the blue **Upload** button.

Creating a custom compute image from an object

1. Authenticate with OCI and navigate to the **Infrastructure Console**.
2. Use the navigation menu to reach the **Compute Custom Images, Core Infrastructure**, page (Hamburger Menu > Compute > Custom Images).
3. Remember to set your **List Scope Compartment**. This will depend on your organization's strategy for managing OCI resources.
4. Click the blue **Image Import** button.
5. In the **Import Image** modal window, select a suitable compartment in the **Create In Compartment** field that conforms to your organization's strategy on managing OCI resources.
6. In the **Name** field enter a unique name to identify the Custom Compute Image. You may want to use the same, resulting name of the image object from the previous step, Upload the Delphix Engine Image as an Object.
7. For **Operating System** select **Linux**.
8. Next, identify an object by specifying its Compartment, Bucket, and Object Name. Or, specify an Object Storage URL. **Note:** The Object Details will identify this value as **URL Path (URI)**.
9. For **Image Type** select **QCOW2**.
10. For **Launch Mode** select **Paravirtualized Mode**.
11. For organizations that have a tagging policy for cloud-based resources, expand the **Tagging Options** section, and define tags.
12. Click the blue **Import Image** button.

Launching the Delphix Engine

1. Authenticate with OCI and navigate to the **Infrastructure Console**.
2. Use the navigation menu to reach the **Compute Instances, Core Infrastructure**, page (Hamburger Menu > Compute > Instances).
3. Remember to set your **List Scope Compartment**. This will depend on your organization's strategy for managing OCI resources.
4. Click the blue **Create Instance** button.
5. In the **Create Compute Instance** window pane, specify a unique name for the VM.
6. For the **Create In Compartment** field, select a suitable compartment that conforms to your organization's strategy on managing OCI resources.
7. In the **Image or operating system** section, click the Change Image button. Switch to the Custom Images tab. Find the Delphix image that corresponds to the instance you wish to deploy. Click the blue **Select Image** button. **Note:** If the Delphix Custom Image is not visible, look for the **Change Compartment** option near the top of the current window pane.
8. Each Availability Domain has its own quota, it is ok to use AD-1, AD-2, or AD-3 - but, be sure to make note of which Availability Domain you are using. **Note:** Compute Instances and attached Storage will need to be in the same Availability Domain.
9. In the **Shape** section click the **Change Shape** button. For **Instance type** specify **Virtual Machine** and for **Shape series** use **Intel Skylake**. Then select an OCI Shape that is supported by Delphix.
10. Continue on to the **Configure networking** section. This part is critical, if the network isn't specified correctly, you are likely to run into firewall issues; please consult your IT or DevOps teams. If your organization is using Network Security Groups (NSGs), mark the **Use Network Security Groups to Control Traffic** checkbox; again, please consult your IT or DevOps teams. Lastly, select the Do Not Assign a Public IP Address radio button; if you must deviate from this guidance then you are highly encouraged to engage your IT or DevOps teams.
11. You may skip the **Boot Volume** section.
12. In the **Add SSH Keys** select the **No SSH Keys** radio option. The Delphix product is a closed appliance and manages users independently.
13. In general, you can skip all of the Advanced Options. For organizations that have a tagging policy for cloud-based resources, expand into the Advanced Management section, and look for the Tagging sub-section to define tags.
14. Click the blue **Create button** - wait about 2-5 minutes for the Delphix Engine instance to boot.

Create block storage volumes

1. Authenticate with OCI and navigate to the **Infrastructure Console**.
2. Use the navigation menu to reach the **Block Volumes, Core Infrastructure**, page (Hamburger Menu > Block Storage > Block Volumes).
3. Remember to set your **List Scope Compartment**. This will depend on your organization's strategy for managing OCI resources.
4. Click the blue **Create Block Volume** button.
5. In the **Create Block Volume** modal window, specify a unique name for this Block Volume. It can be helpful if this name is descriptive or identifies the VM it is intended to be attached to and ends in a sequence number.
6. For the **Availability Domain**, this value **MUST** be the same Availability Domain used for the Delphix Engine instance, otherwise, this volume will not be available for use.
7. In the **Volume Size and Performance** section, select the **Custom** option. Set the size of the volume to be sufficiently large, with room for growth, to support the databases that will be virtualized, or masked, by this Delphix Engine. And, set the **Default Volume Performance** to the **Higher Performance** setting.
8. A **Backup Policy** is not required and can be left blank or **No Backup Policy Selected**. However, depending on your organization's needs, you may consider selecting a Backup Policy.
9. For **Encryption**, it is ok to use the default option, **Encrypt Using Oracle-Managed Keys**. Optionally, if you want, or need, to manage encryption keys independently then use the Encrypt Using Customer-Managed Keys option.

10. For organizations that have a tagging policy for cloud-based resources, expand the **Tagging Options** section, and define tags.
11. Uncheck the checkbox that says **View Detail Page After This Block Volume is Created**. This will prevent you from navigating away from the Block Volumes page, because, more often than not, you will need to create multiple Block Volumes at the same time.
12. Click the blue **Create Block Volume** button.
13. A Delphix Engine requires a minimum of four (4) equally sized Block Volumes. Repeat Steps 4-12 as many times as necessary.

Attach block storage volumes

1. Authenticate with OCI and navigate to the **Infrastructure Console**.
2. Use the navigation menu to reach the **Block Volumes, Core Infrastructure**, page (Hamburger Menu > Block Storage > Block Volumes).
3. Remember to set your **List Scope Compartment**. This will depend on your organization's strategy for managing OCI resources.
4. From the list of pre-existing Block Volumes, identify the resources you wish to attach to a Delphix Engine and wait until the volume's state becomes Available.
5. Select one of the **Block Volumes** to enter the **Block Volume Details** page.
6. On the left-hand side, locate the **Resources** menu and select **Attached Instances**.
7. If the Block Volume has not been previously attached to another VM, then you will be able to click the blue **Attach to Instance** button.
8. In the Attach to Instance modal window, specify the **Attachment Type** as **Paravirtualized**. Currently, iSCSI is not supported.
9. For **Access Type** use the **READ/WRITE** option.
10. Next, identify a Delphix Engine by selecting an instance, or by specifying an instance OCID. If you don't see the Delphix Engine instance in the Select an Instance drop-down menu, you may need to use the Change Compartment option. Block Volumes can only be attached to VM instances that were created in the same Availability Domain - if these values do not match, you will need to either re-provision Block Volumes or the Delphix Engine, in the correct Availability Domain.
11. Click the blue Attach button.
12. Repeat Steps 4-11 until all associated Block Volume resources have been attached to the Delphix Engine instance.

Configuring masking

Once deployed, go to [First Time Setup](#) section to learn how to activate the masking service now that you have the software installed.

VMware installation

The Delphix Engine is a virtual appliance that runs on a hypervisor. In this section, you'll find requirements to run Delphix on VMware including supported versions and instance configurations as well as recommended configuration parameters for optimal performance.

The Delphix Engine is intensive both from a network and a storage perspective. If the Delphix Engine competes with other virtual machines on the same host for resources it will result in increased latency for all operations. As such, it is crucial that your ESXi host is not over-subscribed, as this eliminates the possibility of a lack of resources for the Delphix Engine. This includes allowing a percentage of CPU resources for the hypervisor itself as it can de-schedule an entire VM if the hypervisor is needed for managing IO or compute resources.

Supported ESX versions

Requirements	Notes
VMware Cloud VMware ESX/ESXi 8.0 VMware ESX/ESXi 7.0, 7.0u1, 7.0u2, 7.0 U3c ESX/ESXi 6.7 U3 VMware ESX/ESXi 6.5 U1, 6.5 U3 VMware ESX/ESXi 6.0	More recent versions of VMware are preferred, such as ESX/ESXi 6.0 - 7.0 U3c

 If a minor release version is listed as supported, then all patch releases applicable to that minor release are certified.

Virtual CPUs

Requirements	Notes
8 vCPUs	<ul style="list-style-type: none"> - CPU resource shortfalls can occur both on an over-committed host as well as competition for host resources during high IO utilization. - CPU reservations are strongly recommended for the Delphix VM so that Delphix is guaranteed the full complement of vCPUs even when resources are overcommitted. - It is suggested to use a single core per socket unless there are specific requirements for other VMs on the same ESXi host.
Never allocate all available physical CPUs to virtual machines	<ul style="list-style-type: none"> - CPU for the ESXi Server to perform hypervisor activities must be set aside before assigning vCPUs to Delphix and other VMs. - We recommend that a minimum of 8-10% of the CPUs available are reserved for hypervisor operation. (e.g. 12 vCPUs on a 128 vCore system).

Memory

Requirements	Notes
128 or higher GB vRAM (recommended) 64GB vRAM (minimum)	<ul style="list-style-type: none"> - The masking service on the Delphix Engine uses its memory to process database and file blocks. - More memory can sometimes improve performance. Memory reservation is a requirement for the Delphix VM. - Overcommitting memory resources in the ESX server will significantly impact the performance of the Delphix Engine. - Reservation ensures that the Delphix Engine will not stall while waiting for the ESX server to page in the engine's memory.

 Do not allocate all memory to the Delphix VM.

Never allocate all available physical memory to the Delphix VM. You must set aside memory for the ESX Server to perform hypervisor activities before you assign memory to Delphix and other VMs. The default ESX minimum free memory requirement is 6% of the total RAM. When free memory falls below 6%, ESX starts swapping out the Delphix guest OS. We recommend leaving about 8-10% free to avoid swapping

For example, when running on an ESX Host with 512GB of physical memory, allocate no more than 470GB (92%) to the Delphix VM (and all other VMs on that host).

Network

Requirements	Notes
The ova is pre-configured to use one virtual ethernet adapter of type VMXNET 3.	<ul style="list-style-type: none"> - Jumbo frames are highly recommended to reduce CPU utilization, decrease latency, and increase network throughput. (typically 10-20% throughput improvement) - If additional virtual network adapters are desired, they should also be of type VMXNET 3.
A 10GbE NIC in the ESX Server is recommended.	For VMs having only gigabit networks, it is possible to aggregate several physical 1GbE NICs together to increase network bandwidth (but not necessarily to reduce latency). Refer to the VMware Knowledge Base article NIC Teaming in ESXi and ESX . However, it is not recommended to aggregate NICs in the Delphix Engine VM.

Storage

 Always attach a minimum of 2 storage pools to the Delphix Engine; one for rpool and the other for domain0 pool.

There are three types of data that Delphix stores on disk, which are:

1. **Delphix VM configuration storage:** stores data related to the configuration of the Delphix VM. VM Configuration Storage includes the VMware ESX configuration data as well as log files.
2. **Delphix Engine system disk storage:** stores data related to the Delphix Engine system data, such as the Delphix .ova settings.

3. **Metadata storage:** stores metadata used by the Masking service.

General requirements

Requirements	Notes
Storage used for Delphix must be provisioned from storage that provides data protection.	For example, using RAID levels with data protection features, or equivalent technology. The Delphix engine product does not protect against data loss originating at the hypervisor or SAN layers.

Delphix VM configuration storage

The Delphix VM configuration should be stored in a VMFS volume (often called a "datastore").

Requirements	Notes
The VMFS volume should have enough available space to hold all ESX configuration and log files associated with the Delphix Engine.	If a memory reservation is not enabled for the Delphix Engine (in violation of memory requirements stated above), then space for a paging area equal to the Delphix Engine's VM memory must be added to the VMFS volume containing the Delphix VM configuration data.

Delphix Engine system disk storage

The VMFS volume must be located on shared storage in order to use vMotion and HA features.

Requirements	Notes
The Delphix Engine system disk should be stored in a VMDK.	The VMDK for the Delphix Engine System Disk Storage is often created in the same VMFS volume as the Delphix VM definition. In that case, the datastore must have sufficient space to hold the Delphix VM Configuration, the VMDK for the system disk, and a paging area if a memory reservation was not enabled for the Delphix Engine.
The Delphix .ova file is configured for a 127GB system drive.	The VMFS volume where the .ova is deployed should, therefore, have at least 127GB of free space prior to deploying the .ova.

Delphix Engine metadata storage

Shared storage is required in order to use vMotion and HA features. In addition to making sure the latest VMware patches have been applied, check with your hardware vendor for updates specific to your hardware configuration. VMDKs (Virtual Machine Disks) or RDMs (Raw Device Mappings) operating in virtual compatibility mode can be used for data storage.

Requirements	Notes
The minimum recommended storage size is 50 GB.	

In addition to making sure the latest VMware patches have been applied, check with your hardware vendor for updates specific to your hardware configuration.

Additional VMware configuration notes

- Running Delphix inside of vSphere is supported.
- Using vMotion on a Delphix VM is supported.
- Device passthrough is not supported.

Installing OVA on VMware

1. Download the OVA file from Delphix’s Download site. Note, you will need a support login from your sales team or a welcome letter. Navigate to “Virtual Appliance” and download the appropriate OVA. If unsure, use the HWv11 OVA type.
2. Login using the vSphere client to the vSphere server (or vCenter Server) where you want to install the Delphix Engine.
3. In the vSphere Client, click **File**.
4. Select **Deploy OVA Template** and then browse to the OVA file. Click **Next**.
5. Select a hostname for the Delphix Engine. This hostname will be used in configuring the Delphix Engine network.
6. Select the data center where the Delphix Engine will be located.
7. Select the cluster and the ESX host.
8. Select one (1) data store for the Delphix OS. This datastore can be thin-provisioned and must have 127GB of free space to accommodate the Delphix operating system.
9. The Delphix VM Configuration Storage requires a minimum of 50GB. The VMFS volume should have enough available space to hold all ESX configuration and log files associated with the Delphix Engine. The Delphix Engine system disk should be stored in a VMDK system drive. The VMFS volume must be located on shared storage in order to use vMotion and HA features.
10. Select the virtual network you want to use. If using multiple physical NICs for link aggregation, you must use vSphere NIC teaming. Do not add multiple virtual NICs to the Delphix Engine itself. The Delphix Engine should use a single virtual network.
11. Click **Finish**. The installation will begin and the Delphix Engine will be created in the location you specified.
12. Once the Delphix Engine has been created proceed to [Setting up the Delphix Engine](#) to configure the system.

Naming requirements

This section describes the naming requirements for Masking Engine objects which are allowed to be created/renamed manually.

Affected configurable objects

configurable objects
algorithm
application
connector
domain
environment
file format
job
profiling group
record type
role
rule set
search expression

For all of the above:

- Leading/trailing white space is not allowed
- The following special characters are not allowed:

Symbol	Name
[open bracket
]	close bracket

Symbol	Name
(open parenthesis
)	close parenthesis
{	open brace
}	close brace
~	tilde
!	exclamation mark
@	at
#	pound
\$	dollar
%	percent
^	carat
*	asterisk
"	quote
?	question mark
:	colon
;	semi-colon
,	comma
/	forward slash
\	back slash

Symbol	Name
\\	double back slash
`	back quote
+	plus
=	equal
<	less than
>	greater than
'	single quote
	pipe

Upgrade

During an upgrade of a Masking Engine to a 6.0 or later release, a name with leading or trailing white space will be automatically trimmed, and a counter value might be appended to the end of the name to prevent a naming conflict. For example:

pre-upgrade name	post-upgrade name	upgrade change
"alg_SecureLookup"	"alg_SecureLookup"	no change
" alg_SecureLookup"	"alg_SecureLookup1"	leading white space trimmed and counter value appended
"alg_SecureLookup "	"alg_SecureLookup2"	trailing white space trimmed and counter value appended

If any name from the above-mentioned "configurable entities" table has a restricted special character - an upgrade will fail with the corresponding error message.

Create/Rename

If an attempt is made to create a new entity (or to modify the name of the existing one) with leading or trailing white space or any of the special characters listed above, the operation will fail on a 6.0 or later release with a corresponding error message.

Environment Export - Import

If any entity name exported from a pre-6.0 version contains leading or trailing white spaces or the special characters listed above, the import operation will fail on a 6.0 or later release with a corresponding error message.

Sync

If a sync bundle from a pre-6.0 version contains leading or trailing white space or any of the special characters listed above, then the Sync import operation will fail on a 6.0 or later release, with a corresponding error message.

Users and roles

The Delphix Masking Service has a flexible and robust users and roles system that allows you to give users fine-grain privileges over what environments they have access to and what tasks they can and can not perform.

What are roles?

A defined role is what is used to give certain users privileges over certain environments and tasks. Roles can be defined by selecting a subset of actions that can be taken on certain objects.

Actions

When defining a role, you can select one or more of the following actions for the role to be able to perform:

- **View:** Be able to view the object and important information about the object.
- **Add:** Be able to add an instance of an object.
- **Update:** Be able to update/edit an instance of an object.
- **Delete:** Be able to delete an instance of an object.
- **Copy:** Be able to create a copy of an object.
- **Export:** Be able to export an object from a Delphix Engine.
- **Import:** Be able to import an exported object into a Delphix Engine

Please note that not all of these actions are available for all objects in the masking service.

Objects

When defining a role, permission to perform the above actions can be defined on a per-object basis. These objects include:

General	Jobs	Settings	Report	Approval Workflow
Environment	Profile Job	Domains	Inventory Report	Approve Inventories
Connection	Masking Job	Algorithms		
Ruleset	Tokenize Job	Plugins		
Inventory	Re-identify Job	Profile Expression		
		Profiler Set		
		File Format		
		JDBC Drivers		

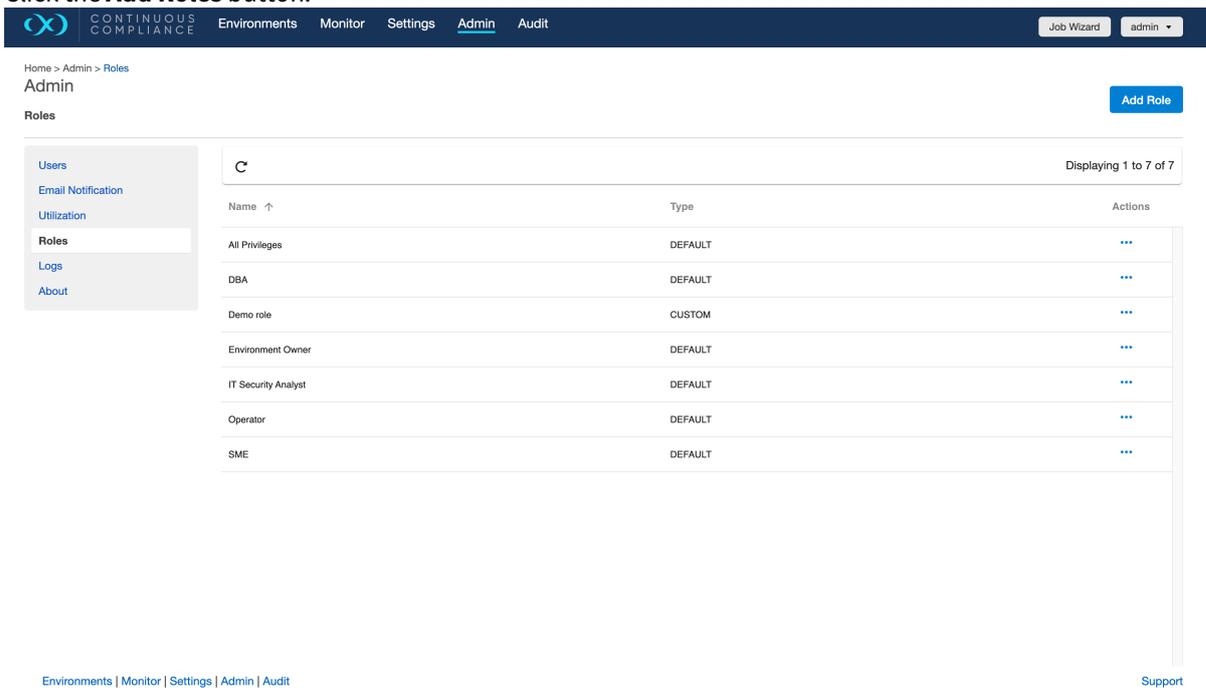
General	Jobs	Settings	Report	Approval Workflow
		Password Vault		
		User		
		Diagnostic		

Refer to [Delphix Masking Terminology](#) for definitions of these objects.
 View privilege for Plugins and JDBC Drivers should be always true for any type of role.
 Environment Export privilege permission is no longer supported.

Adding a role

To add a role follow these steps:

1. Login into the **Masking Engine**, select the **Admin** tab and select **Roles** from the left navigation panel.
2. Click the **Add Roles** button.



3. A full-screen dialog will appear for adding a new role. Enter a **Role Name**. The far-left column lists the items for which you can set privileges.

Add Role ✕

Role Name

General	View	Add	Update	Delete	Copy	Import	Export
Environment	<input type="checkbox"/>						
Connection	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Rule Set	<input type="checkbox"/>						
Inventory	<input type="checkbox"/>		<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>
Jobs	View	Add	Update	Delete	Run		
Profile Job	<input type="checkbox"/>						
Masking Job	<input type="checkbox"/>						
Tokenize Job	<input type="checkbox"/>						
Re-identify Job	<input type="checkbox"/>						
Settings	View	Add	Update	Delete			
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Report	View						
	<input type="checkbox"/>						
Approval Workflow	View						
	<input type="checkbox"/>						

Cancel Save

4. Select the checkboxes for the corresponding privileges that you want to apply. If there is no checkbox, that privilege is not available. For example, if you want this role to have View, Add, Update, and Run privileges for masking jobs, select the corresponding checkboxes in the **Masking Job** row.

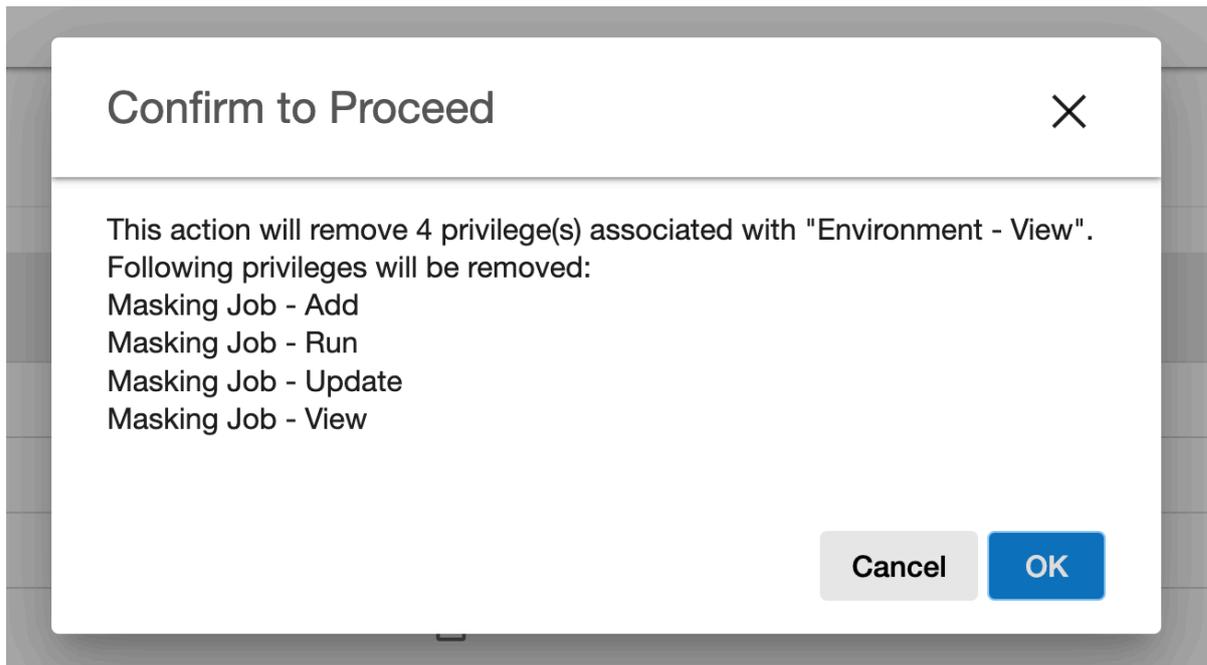
Add Role ✕

Role Name
Demo role

General	View	Add	Update	Delete	Copy	Import	Export
Environment	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Connection	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Rule Set	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Inventory	<input type="checkbox"/>		<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>
Jobs	View	Add	Update	Delete	Run		
Profile Job	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Masking Job	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>		
Tokenize Job	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Re-identify Job	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Settings	View	Add	Update	Delete			
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Report	View						
	<input type="checkbox"/>						
Approval Workflow	View						
	<input type="checkbox"/>						

Cancel Save

5. If you are removing any privileges with dependents selected then, a dialog box will appear asking for confirmation for removing privileges with all its dependents. For example, Masking Job [View, Add, Update, Run] depends on Environment - View. So if the user tries to remove Environment - View then it will ask for confirmation to remove it with the list of dependents. If users do not wish to remove it then they can select the **Cancel** or **Close icon** else can select **OK** and move forward with changes.



6. When you are finished assigning privileges for this Role, click **Save**.

Role Type

With the 13.0.0.0 release, a New parameter Role type has been added to the product. There are 2 types of roles. **CUSTOM** and **DEFAULT**. All built-in system roles will have a type of **DEFAULT** and others will be the type of **CUSTOM**.

There have been 5 roles added to the product with the role type **DEFAULT**.

1. IT Security Analyst
2. DBA
3. SME
4. Operator
5. Environment Owner

If a role with the name 'All Privileges' exists on the product then it will be marked as a **DEFAULT** role type.

If a role with any of the above names already exists then the role being added will be renamed by appending sequence number. For example, If '**DBA**', and '**DBA 1**' are already present in the system then the role with the name '**DBA 2**' will be added to the product with **DEFAULT** role type.

Recommended roles

While every organization will differ in what users and roles they define, Delphix uses these common/popular roles. Please note that each defined user can only have one role assigned to them.

Administrator — This role is assigned by enabling a user's Administrator setting in either the UI or API. A user with this role has unrestricted access to all the engine functions. Specifically, the user has all privileges available through the roles system and the following additional, Administrator-only privileges:

- [Sync](#)
- A User's `apiAccess` and `userStatus` setting
- Audit Page

- Admin > Users Generate Key Button
- Admin > Email Notification
- Admin > Utilization
- Deletion of any object: An Admin can delete any object, such as any Algorithm, Domain, Profile Expression, or Profile Set. In contrast, a user with the **All Privileges** role can only delete objects they created.
- Settings > Roles

IT Security analyst — Unrestricted access for all settings functions; access to all application functions except environment and environment create, delete, update.

All Privileges — Unrestricted access to an application environment; central admin or security analyst will determine if this role can modify settings.

DBA — Manage connections for the application database, scripting, and scheduling (no settings).

SME/Analyst/Developer — Manage inventories, create, and view jobs.

Operator — All job privileges.

Environment Owner — Approve workflow and inventories, privileges to view for settings and environment.

Modifying Roles

Users can perform 4 types of action on roles by selecting action options from the far right of the grid.

The screenshot shows the 'Admin' section of the Continuous Compliance interface. A navigation bar at the top includes 'Environments', 'Monitor', 'Settings', 'Admin', and 'Audit'. Below the navigation bar, there is a breadcrumb trail 'Home > Admin > Roles' and a sub-header 'Admin'. A sidebar on the left contains links for 'Users', 'Email Notification', 'Utilization', 'Roles', 'Logs', and 'About'. The main content area displays a table of roles with columns for 'Name', 'Type', and 'Actions'. A context menu is open over the 'Demo role' entry, showing options for 'View', 'Edit', 'Duplicate', and 'Delete'. The table also includes a search bar and a 'Displaying 1 to 7 of 7' indicator.

Name	Type	Actions
All Privileges	DEFAULT	...
DBA	DEFAULT	...
Demo role	CUSTOM	...
Environment Owner	DEFAULT	...
IT Security Analyst	DEFAULT	...
Operator	DEFAULT	...
SME	DEFAULT	...

1. **View** - The user can view the role details. Everything on the dialog form will be disabled in case of view is selected.

View Role



Role Name
Demo role

General	View	Add	Update	Delete	Copy	Import	Export
Environment	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Connection	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Rule Set	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Inventory	<input type="checkbox"/>		<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>
Jobs	View	Add	Update	Delete	Run		
Profile Job	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Masking Job	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Tokenize Job	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Re-identify Job	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Settings	View	Add	Update	Delete			
Domains	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Algorithms	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Plugins	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Profile Expression	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			

Close

2. **Edit** - The user can edit the privileges of roles. The role name is not allowed to be edited. It will be disabled.

Edit Role



Role Name
Demo role

General	View	Add	Update	Delete	Copy	Import	Export
Environment	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Connection	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Rule Set	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Inventory	<input type="checkbox"/>		<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>
Jobs	View	Add	Update	Delete	Run		
Profile Job	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Masking Job	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Tokenize Job	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Re-identify Job	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Settings	View	Add	Update	Delete			
Domains	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Algorithms	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Plugins	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Profile Expression	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			

Cancel Save

3. **Duplicate** - On selecting a duplicate option, privileges will be pre-selected to the new role screen dialog, and the user can give a new name and duplicate the role.

Duplicate Role

✕

Role Name

^ General	View	Add	Update	Delete	Copy	Import	Export
Environment	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Connection	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Rule Set	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Inventory	<input type="checkbox"/>		<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>
^ Jobs	View	Add	Update	Delete	Run		
Profile Job	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Masking Job	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Tokenize Job	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Re-identify Job	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
^ Settings	View	Add	Update	Delete			
Domains	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Algorithms	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Plugins	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Profile Expression	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			

Cancel

Save

4. Delete

Delete Role

Are you sure you want to delete "Demo role" role?

Cancel

Confirm

Role with **CUSTOM** type can be added, edited, deleted, duplicated, and viewed.

Role with **DEFAULT** type is not allowed to be added, edited, or deleted. It can be only viewed and duplicated.

To Modify or Add a role using Masking API, follow these steps:

1. Access the API client on your Masking Engine, from <http://myMaskingEngine.myDomain.com/masking/api-client>.
2. Login into the Masking Engine and select the Role endpoint.

role		^
GET	/roles Get all roles	⌵ 🔒
POST	/roles Create role	⌵ 🔒
GET	/roles/{roleId} Get role by ID	⌵ 🔒
PUT	/roles/{roleId} Update role by ID	⌵ 🔒
DELETE	/roles/{roleId} Delete role by ID	⌵ 🔒

- Execute API requests by providing request parameters and body as mentioned in the example. ([Sample JSON for Add/Update role API](#))

What are users?

Once you have your roles defined, it is time to create users with those roles. We highly recommend creating independent users for each individual who will have access to the masking service.

Adding a user

To create a new user using the Masking UI follow these steps:

- Login into the **Masking Engine** and select the **Admin** tab.
- Click **Add User** at the upper right of the Users screen.
- You will be prompted for the following information:
 - First Name** — (Optional) The user's given name
 - Last Name** — (Optional) The user's surname
 - User Name** — The login name for the user
 - Email** — The user's e-mail address (mailable from the Delphix Masking Engine server for purposes of job completion e-mail messages)
 - Password** — The password that the Delphix Masking Engine uses to authenticate the user on the login page. The password must be at least six characters long but no longer than 12 characters, and contain a minimum of one uppercase character, one wild character (!@#\$%^&*), and one number.
 - Confirm Password** — Confirm the password with double-entry to avoid data entry errors.
 - Administrator** — (Optional) Select the Administrator checkbox if you want to give this user Administrator privileges. (Administrator privileges allow the user to perform all Delphix Masking Engine tasks, including creating and editing users in the Delphix Masking Engine.) If you select the Administrator checkbox, the Roles and Environments fields disappear because Administrator privileges include all roles and environments.
 - Role** — Select the role to grant to this user. The choices here depend on the custom roles that you have created. You can assign one role per user name.
 - Environment** — Enter as many environments as this user will be able to access. Granting a user access to a given environment does not give them unlimited access to that environment. The user's access is still limited to their assigned role.

Add User

First Name	Last Name
<input style="width: 95%;" type="text"/>	<input style="width: 95%;" type="text"/>
User Name	Email
<input style="width: 95%;" type="text"/>	<input style="width: 95%;" type="text"/>
Password	Confirm Password
<input style="width: 95%;" type="password"/>	<input style="width: 95%;" type="password"/>
Administrator	<input type="checkbox"/>
Role	
<input style="width: 95%;" type="text" value="Choose Role"/>	
Environment	
<input style="width: 95%;" type="text" value="Select Environment Name"/>	
<input type="button" value="Cancel"/> <input type="button" value="Save"/>	

4. When you are finished, click **Save**.

When a user is created, it's Account Status is *Active* by default.

To create a new user using the Masking API follow these steps:

1. Access the API client on your Masking Engine, from <http://myMaskingEngine.myDomain.com/masking/api-client>.
2. Login into the Masking Engine and select the User endpoint.
- 3.

user		Show/Hide List Operations Expand Operations
GET /users		Get all users
POST /users		Create user
DELETE /users/{userId}		Delete user by ID
GET /users/{userId}		Get user by ID
PUT /users/{userId}		Update user by ID
POST /users/forgot-password		Send Reset password mail to the user
POST /users/reset-password		Reset new password for the user

Click Create users at the upper right of /users section and refer to the Example Value for parameters required for new users.

POST /users Create user

Response Class (Status 201)
Success

Model **Example Value**

```
{
  "userName": "DelphixUser1",
  "password": "Password_123",
  "firstName": "First",
  "lastName": "Last",
  "email": "user@delphix.com",
  "isAdmin": false,
  "showWelcome": true,
  "userStatus": "ACTIVE",
  "nonAdminProperties": {
    "roleId": 1,

```

Response Content Type:

Parameters

Parameter	Value	Description	Parameter Type	Data Type
body	<pre>{ "userName": "DelphixUser1", "password": "Password_123", "firstName": "First", "lastName": "Last", "email": "user@delphix.com", "isAdmin": false, </pre>	The user to create	body	Model Example Value

Parameter content type:

Response Messages

HTTP Status Code	Reason	Response Model	Headers
400	Bad request		
401	Unauthorized access		
404	Not found		
409	Conflict		

[Hide Response](#)

Curl

4. Enter valid User creation JSON in the **body** section, refer to sample create users JSON ([Sample New User Create JSON](#))
5. Click on **Execute API Request**.

Updating a user

To update user information using Masking UI, follow these steps:

1. Login into the **Masking Engine** and select the **Admin** tab.
2. Select the **Edit** icon next to the user you want to edit. The **Edit User** screen will appear with existing user details.
3. The following user information can be modified through the **Edit User** screen:
 - First Name
 - Last Name
 - Email Address
 - Password
 - Administrator Status
 - Welcome Page Status
 - Account Status (cannot be changed to *Locked*)
 - User Roles (non admin users only)
 - User Environments (non admin users only)

Edit User

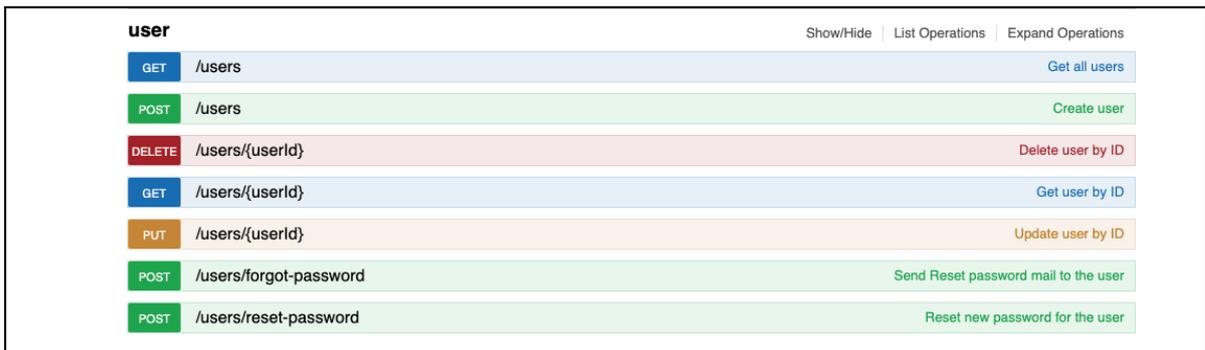
First Name	Last Name
<input type="text"/>	<input type="text"/>
User Name	Email
<input type="text" value="admin"/>	<input type="text" value="abcd@delphix.com"/>
Change Password	<input checked="" type="checkbox"/>
Password	Confirm Password
<input type="text"/>	<input type="text"/>
Administrator	<input checked="" type="checkbox"/>
Enable Welcome Page	<input checked="" type="checkbox"/>
Enable Notification Popup	<input type="checkbox"/>
Account Status	<input type="text" value="Active"/>

4. When you are finished, click **Save**.

 User's Account Status will be automatically changed to *Locked* on multiple invalid login attempts.

To update user information using Masking API, follow these steps:

1. Access the API client on your Masking Engine, from <http://myMaskingEngine.myDomain.com/masking/api-client>.
2. Login into the Masking Engine and select the User endpoint.

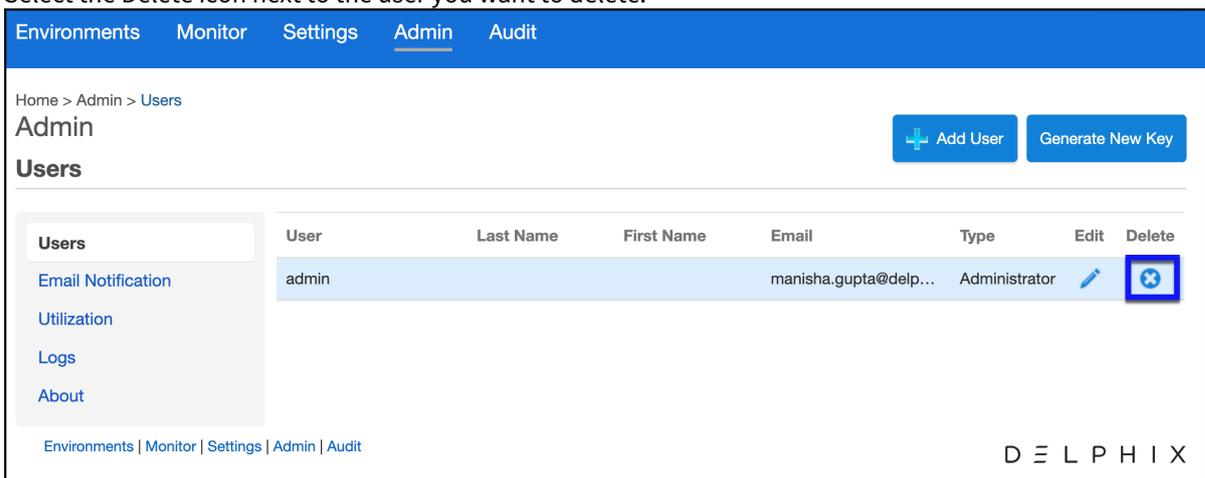


3. click **Update user by ID** at the upper right of the section and refer to the **Example Value** for parameters required for new users.
4. Enter valid User creation JSON in the **body** section, refer to sample create users JSON. ([Sample User JSON](#))
5. Click on **Execute API Request**.

Deleting a user

To delete a user using the Masking UI follow these steps:

1. Login into the **Masking Engine** and select the **Admin** tab.
2. Select the Delete icon next to the user you want to delete.



3. In the confirmation box select **OK**.

To delete a user using Masking API follow these steps:

1. Access the API client on your Masking Engine, from the `http://myMaskingEngine.myDomain.com/masking/api-client`.
2. Login into the Masking Engine and select the User endpoint.

user Show/Hide | List Operations | Expand Operations

GET	/users	Get all users
POST	/users	Create user
DELETE	/users/{userId}	Delete user by ID
GET	/users/{userId}	Get user by ID
PUT	/users/{userId}	Update user by ID
POST	/users/forgot-password	Send Reset password mail to the user
POST	/users/reset-password	Reset new password for the user

3. Click Delete a user by ID at the upper right of /users section.

DELETE /users/{userId} Delete user by ID

Parameters ⓘ

Parameter	Value	Description	Parameter Type	Data Type
userId	<input type="text" value="(required)"/>	The ID of the user to delete	path	integer

Response Messages

HTTP Status Code	Reason	Response Model	Headers
200	Success		
400	Bad request		
401	Unauthorized access		
404	Not found		

4. Enter the **userId** for the user to be deleted
5. Click on **Execute API Request**

Sample JSON

This section contains the following articles:

- [All privileges](#)
- [Create new user](#)
- [User update](#)

All privileges

```
{
  "roleName": "All Privileges",
  "environment": {
    "copy": true,
    "create": true,
    "delete": true,
    "export": true,
    "import": false,
    "run": false,
    "update": true,
    "view": true
  },
  "connector": {
    "copy": false,
    "create": true,
    "delete": true,
    "export": false,
    "import": false,
    "run": false,
    "update": true,
    "view": true
  },
  "ruleset": {
    "copy": true,
    "create": true,
    "delete": true,
    "export": false,
    "import": false,
    "run": false,
    "update": true,
    "view": true
  },
  "inventory": {
    "copy": false,
    "create": false,
    "delete": false,
    "export": true,
    "import": true,
    "run": false,
    "update": true,
    "view": true
  },
  "profileJob": {
    "copy": false,
    "create": true,
    "delete": true,
    "export": false,
    "import": false,
```

```
"run": true,
"update": true,
"view": true
},
"maskingJob": {
  "copy": false,
  "create": true,
  "delete": true,
  "export": false,
  "import": false,
  "run": true,
  "update": true,
  "view": true
},
"tokenizeJob": {
  "copy": true,
  "create": true,
  "delete": true,
  "export": true,
  "import": true,
  "run": true,
  "update": true,
  "view": true
},
"reidentifyJob": {
  "copy": true,
  "create": true,
  "delete": true,
  "export": true,
  "import": true,
  "run": true,
  "update": true,
  "view": true
},
"scheduler": {
  "copy": false,
  "create": true,
  "delete": true,
  "export": false,
  "import": false,
  "run": false,
  "update": true,
  "view": true
},
"domain": {
  "copy": false,
  "create": true,
  "delete": true,
  "export": false,
  "import": false,
  "run": false,
  "update": true,
```

```
    "view": true
  },
  "algorithm": {
    "copy": false,
    "create": true,
    "delete": true,
    "export": false,
    "import": false,
    "run": false,
    "update": true,
    "view": true
  },
  "jdbcDriver": {
    "copy": false,
    "create": true,
    "delete": true,
    "export": false,
    "import": false,
    "run": false,
    "update": true,
    "view": true
  },
  "passwordVault": {
    "copy": false,
    "create": false,
    "delete": false,
    "export": false,
    "import": false,
    "run": false,
    "update": false,
    "view": false
  },
  "plugin": {
    "copy": false,
    "create": true,
    "delete": true,
    "export": false,
    "import": false,
    "run": false,
    "update": true,
    "view": true
  },
  "profileExpression": {
    "copy": false,
    "create": true,
    "delete": true,
    "export": false,
    "import": false,
    "run": false,
    "update": true,
    "view": true
  },
}
```

```
"profileSet": {
  "copy": false,
  "create": true,
  "delete": true,
  "export": false,
  "import": false,
  "run": false,
  "update": true,
  "view": true
},
"fileFormat": {
  "copy": false,
  "create": true,
  "delete": true,
  "export": false,
  "import": false,
  "run": false,
  "update": false,
  "view": true
},
"user": {
  "copy": false,
  "create": true,
  "delete": true,
  "export": false,
  "import": false,
  "run": false,
  "update": true,
  "view": true
}
}
```

Create new user

```
{
  "userName": "DelphixUser1",
  "password": "Password_123",
  "firstName": "First",
  "lastName": "Last",
  "email": "user@delphix.com",
  "isAdmin": false,
  "showWelcome": true,
  "userStatus": "ACTIVE",
  "nonAdminProperties": {
    "roleId": 1,
    "environmentIds": [
      1
    ]
  }
}
```

User update

```
{
  "userName": "DelphixUser1",
  "password": "Password_123",
  "firstName": "First",
  "lastName": "Last",
  "email": "user@delphix.com",
  "isAdmin": false,
  "showWelcome": true,
  "userStatus": "ACTIVE",
  "nonAdminProperties": {
    "roleId": 1,
    "environmentIds": [
      1
    ]
  }
}
```

Best practices for defining masking roles

Introduction

The Delphix Masking Engine contains a role definition capability that enables admins to easily create roles for users. This section describes the typical roles and privileges that can be granted to users. It is recommended that the masking administrator implementing these roles consult IT Security and follow existing policies for data access. Roles are added by clicking the appropriate checkboxes within the add role function in the Settings tab. A sample RACI document and examples of roles / privileges are located below.

Roles for operating the Delphix Masking Engine are shared primarily between the masking administration team and the teams that support the applications that will be on-boarded to the Masking Engine. The admin will manage central functions of the engine including definition of custom domains, profiler expressions, algorithms, role and user definitions. The masking Engine is flexible enough to enable application teams with these functions as well, but it is recommended that these shared functions be managed by the admin team. The admin team should have an account registered with Delphix Support and be the main interface for issues and maintenance support from Delphix.

Masking processes can be developed for each application by the central admin team or the individual application teams, often determined by the volume of applications to be on-boarded. The RBAC model employed by Delphix Masking can support different implementation models. Your Delphix support team can assist in constructing roles to meet your needs.

Once roles are defined, they can be assigned to individual user IDs for the environments that those users have responsibility. Administrators will have access to all masking settings and environments by default.



1. Administrator access provides unlimited access to all functions and environments; this role should be granted to the central administration team.
2. All privileges is a default role (predefined) which will provide all functions for each environment a user is given access to.
3. Connector access should be controlled and administered by personnel responsible for database access.

Sample RACI

Teams: IT Security DM = Data masking admin team Application = App owner/SME DBA = Database admin QA = QA/ Test environment owner PM = project management

Role	Description	Accountable	Responsible	Consulted	Informed
Security Policy	Determine data types that are sensitive for the enterprise.	IT Security	IT Security	DM, Application	DBA, QA
Program Management	Maintain program plan and implementation schedule, tracking and reporting.	PM	DM, Application	QA, IT Security	DBA

Role	Description	Accountable	Responsible	Consulted	Informed
Inventory Management	Apply security policy to application schemas/ files.	Application	DM, Application	DBA, QA	IT Security
Data Masking	Build, maintain, schedule masking processes.	Application	DM, DBA	QA	IT Security
Masked Data Validation	Review and approve inventories and masked data.	Application	Application, DBA, QA	DM	IT Security
Masked Data Deployment	Deploy masked data to required environments.	Application	Application, DBA, QA	DM, QA	IT Security
Environment Audit	Assure applications are compliant with masking.	IT Security	IT Security	DM, DBQ, QA	Application
Masking Administration	Manage masking tool central functions, create domains, profiler expressions, roles, users.	DM	DM	Application, IT Security, DBA	QA

Sample roles for Masking

Role	Description	<i>*Delphix Masking Functions</i>
Administrator	Manages masking server updates and upgrades; works with IT Security to update domains, algorithms and profiler expressions / sets.	Unrestricted access to all the engine functions. The Admin role is assigned via the checkbox in the add user page of the UI.
IT Security Analyst	Determines domains to be masked and high-level method for each domain and communicates them to administrator for inclusion in masking engine, responsible for masking audit functions.	Unrestricted access for all settings functions; access to all application functions except environment and environment create, delete, update.

Role	Description	*Delphix Masking Functions
Application Roles (per environment)		
All Privileges	Super user for an environment.	Unrestricted access for an application environment; central admin or security analyst will determine if this role can modify settings.
DBA	Manages user privileges, database performance and schema definition.	Manage connectors for application database, scripting and scheduling (no settings).
SME / Analyst / Developer	Application subject matter expert, application developer, data analyst, application architecture.	Manage inventories, create, view jobs.
Operations Roles (per environment)		
Operator	Schedule jobs, execute jobs, verify results, run automation scripts.	All job privileges.
Environment Owner	Determine workflow, monitor tool usage for environment.	Approve workflow and inventories, privileges to view for settings and environment.

Audit logs

Delphix helps you keep a record of user actions taken in the UI or directly through our REST APIs. You can access these audit logs directly from our UI or through our APIs.

Audit log UI page

The Audit Log page can be found in the UI under the Audit tab. This page contains information on what action occurred, the user that performed the action, and the time at which the action occurred. It also provides the ability to filter based on:

- user
- time range
- arbitrary search string
- action type or action target, or both (create, connector or create a database connector)

Audit log APIs

With 5.3.2.0, Delphix introduced an endpoint to get all Audit Logs. This endpoint contains the user name, action type, target, status, start time, and end time. For more information please refer to [API documentation](#).

What gets logged?

User actions are categorized into the following:

Cancel	Create	Delete	Edit	Export	Get	Get All
Import	Lock	Login	Logout	Run	Test	Unlock

The objects that user actions target are categorized into the following:

Algorithm	Analytics	Application	Application Log	Async Task	Audit Log	
Column Metadata	Database Connector	Ruleset Connector	Database Ruleset	Domain	Encryption Key	Environment
Execution	File Connector	File Download	File Field Metadata	File Format	File Metadata	File Ruleset
File Upload	LDAP	Mainframe Dataset Connector	Mainframe Dataset Field Metadata	Mainframe Dataset Format	Mainframe Dataset Metadata	Mainframe Dataset Ruleset

Masking Job	Profile Expression	Profile Job	Profile Set	Re Identification Job	Role	SSH Key
SSO	Syncable Object	System Information	Table Metadata	Tokenization	User	

Retention policy

The default policy stores the last one million Audit Log entries. Any entries older than the most recent million are removed daily. Additionally, there is a fail-safe mechanism that prevents an attacker from forcing an unbounded number of actions to be logged to overload the system's disk space. In the event that such an attack occurs, Delphix also logs it to the application logs.

Recommendation

If a full record of all Audit Log entries is desired, Delphix recommends using the new API to periodically retrieve new entries from the Audit Logs.

Kerberos configuration

Introduction

As of 5.3.0.0, the Continuous Compliance Engine supports Kerberos authentication for Oracle, MS SQL Server, and Sybase connections. Utilizing this service requires the presence of a Kerberos Key Distribution Center (KDC) server as well as additional configuration actions to be done on both the Masking Engine and the database. This document presents configuration instructions for enabling and using Kerberos on the Continuous Compliance Engine, as well as reference configurations for enabling Kerberos on the Databases. Although other configurations are possible, the configurations in this document have been validated by Delphix.

i Kerberos is not supported for containerized masking deployments at this time. This is a roadmap item that is expected to be added for containerized masking at some future point.

Terminology

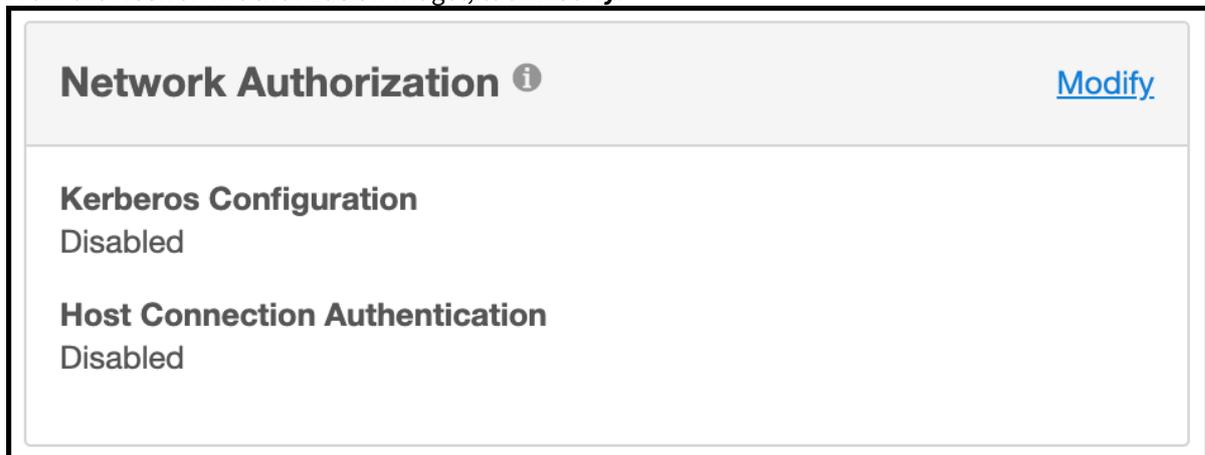
Throughout this document, the following example values are used. To recreate these reference environments, these values must be replaced with real values appropriate for your network environment: - .bar.com - the DNS domain of the network - BAR.COM - the Kerberos domain - me-host - the hostname of the Masking Engine - foo-kcd - the hostname KDC server - krbuser - the Kerberos principal to be granted access to the database for masking

Configuring Kerberos on the appliance

This section details the steps required to configure Kerberos on your appliance.

Launch the Delphix Server Setup UI and perform the following steps to enable Kerberos:

1. From the **Network Authorization** widget, click **Modify**.



2. Select the checkbox before **Use Kerberos authentication to communicate with remote hosts** field.
3. Click the plus symbol to add record(s) for your KDCs, and populate other fields appropriately for your network environment. Upon pressing **Save**, your configuration will be tested. If the engine is able to authenticate to the KDC with the supplied configuration, the configuration is applied immediately.

Network Authorization ✕

KERBEROS CONFIGURATION

Use Kerberos authentication to communicate with remote hosts

Kerberos Key Distribution Center host(s)

+

Hostname [^]	Port
foo-kdc.bar.com	88

Realm

BAR.COM

Principal

krbuser

Keytab

_krbuser_keytab_base64_ ↻

HOST CONNECTION AUTHENTICATION

When connecting to hosts, you can provide username-password pairs when setting up the connection, or you can utilize one or more Enterprise Password Vault systems by adding them to your engine setup.

Click the + to add a vault

+

Vault name	Hostname	Port	Vault Type	Auth Method
------------	----------	------	------------	-------------

Cancel
Save

Creating masking database connectors using Kerberos

Once the Delphix Appliance is configured for Kerberos, creating Connectors using Kerberos authentication is simple:

Create Connection

Type

Database - Oracle Basic Advanced

Connection Name **Port**

Schema Name Use Kerberos Authentication

Host Name/ IP **Principal Name**

SID **Password**

Custom Properties File [?](#)

Assuming you are using the same user principal configured in Server Setup, the keytab will be used and it is unnecessary to enter a password in the Connector definition.

For Sybase database Connectors, it is necessary to supply the service principal name as an additional configuration item. For Oracle DB, this value is determined automatically. For MS SQL Server it is determined based on the reverse DNS mapping of the Server Name (refer to the section on MS SQL Server below).

- ⓘ If any changes are made to the underlying **krb5.conf** configuration file, these changes will not be recognized by the engine until after the next database connection attempt. Therefore, expect to have to hit "Test Connection" twice after making any changes to the **krb5.conf** file. It does not matter if the first connection attempt succeeds or fails.

Reference database configurations

The following is a series of reference Kerberos configuration procedures and troubleshooting notes for the supported databases. These are meant to serve as examples to be further customized according to the user's specific network environment and security needs.

Oracle database

Overview

This document describes how to set up an Oracle DB instance for Kerberized connections. The following steps are described:

- Creating a service principal and adding it to the DB system
- Configuring the database to use Kerberos authentication
- Creating DB users identified via Kerberos
- Troubleshooting tips

Prerequisites

This document assumes you already have a kerberized network environment with an MIT Kerberos KDC. These procedures have been tested successfully with Oracle database versions 11.2.0.2, 11.2.0.4 and 12.2.1. Oracle database version 12.1.0.1 did not work in our testing.

You will need the following from your Kerberos environment: - The krb5.conf file - A user principal and associated password or key tab you'd like to use to log into the database - The ability to create a service principal for the Oracle DB and retrieve the associated key tab.

This section of the document uses these example values in addition to those mentioned above:

- The Oracle database is: [ora-db.bar.com](#).
- The Oracle service name is: oracle

Creating the Oracle Service Principal

The service principal will be named:

Notice that the hostname is whatever the database system thinks its hostname is - that is, the output of "uname -n" on the database system, rather than the actual DNS name of the database system. Typically, these values would be the same, but this is not always the case.

On the KDC, run:

```
# kadmin.local kadmin.local: addprinc -randkey oracle/ora-db@bar.com
kadmin.local: ktadd -norandkey -k /var/tmp/ora-db.keytab oracle/ora-db@bar.com
```

Copy the resulting keytab file (/var/tmp/ora-db.keytab) to the Oracle DB system at this location: /etc/v5srvtab

As root on the Oracle DB system, ensure that the keytab has the correct permissions:

```
# chown root:oinstall /etc/v5srvtab
```

```
# chmod 440 /etc/v5srvtab
```

Finally, this is a good opportunity to copy /etc/krb5.conf from the KDC to /etc/krb5.conf on the Oracle DB system. This file should be readable by all users.

Configuring the Oracle Database for Kerberos

Log into the Oracle DB system as the appropriate user for the database in question.

```
$ cd $ORACLE_HOME $ vi network/admin/sqlnet.ora
```

Add the following for Oracle 11:

```
SQLNET.KERBEROS5_CONF=/etc/krb5.conf
```

```
SQLNET.AUTHENTICATION_SERVICES=(BEQ,KERBEROS5) SQLNET.KERBEROS5_CONF_MIT=true
```

```
SQLNET.AUTHENTICATION_KERBEROS5_SERVICE=oracle
```

Or the following for Oracle 12:

```
NAMES.DIRECTORY_PATH=(TNSNAMES, EZCONNECT, HOSTNAME) SQLNET.KERBEROS5_CONF=/
```

```
etc/krb5.conf SQLNET.AUTHENTICATION_SERVICES=(BEQ,KERBEROS5PRE,KERBEROS5)
```

```
SQLNET.KERBEROS5_CONF_MIT=true SQLNET.AUTHENTICATION_KERBEROS5_SERVICE=oracle
```

If the database is Oracle 11 (not necessary on Oracle 12): `$ vi dbs/init.ora` Add this line at the end:

```
OS_AUTHENT_PREFIX=""
```

Creating a DB User Identified via Kerberos

Log into the Oracle DB system as the appropriate database user and open a database session as the DBA:

```
$ sqlplus / as sysdba
```

On Oracle 12, you may wish to alter your session to create the user in one of the PDBs: `SQL> alter session set container=MYPDB;`

Create the user that will connect to the DB using kerberos:

```
SQL> create user krbdbuser identified externally as 'krbuser@BAR.COM';
```

Grant the user privileges necessary for masking.

This example grants all privileges for the sake of simplicity:

Oracle 11:

```
SQL> grant all privilege to krbdbuser;
```

Oracle 12: (Customize permissions as necessary for your environment).

```
SQL> grant connect,resource to krbdbuser;

SQL> grant create tablespace, drop tablespace to krbdbuser;

SQL> grant create table to krbdbuser;

SQL> grant create sequence to krbdbuser;

SQL> grant select_catalog_role to krbdbuser;

SQL> grant unlimited tablespace to krbdbuser;

SQL> grant select_catalog_role to krbdbuser;

SQL> grant alter system to krbdbuser;

SQL> grant sysoper to krbdbuser;
```

```
SQL> grant dba to krdbuser;
```

Troubleshooting Tips

- Connecting via JDBC with Kerberos authentication from Continuous Compliance involves two steps - a Kerberos login, followed by JDBC connect. A failure stack with an error in the login function indicates a misconfiguration on either the engine or KDC - the engine hasn't even attempted to communicate with the database at that point. Failure stacks are saved in the debugging log for masking.
- Login exceptions that mention a checksum error mean either the password or keytab supplied doesn't match the expected password/key on the KDC for the principal you're trying to use. Server Setup verifies that your keytab works at configuration time, but it could stop working if the key for your principal is updated on the KDC.
- Prior to version 12, Oracle databases instances assume they can create/write a particular temporary file to store Kerberos credentials for the DB. This means if you attempt to run multiple kerberized instances of Oracle 11 on the same system or VM, and the databases run as different system users, the first Oracle instance that performs Kerberos auth will create and own this file. Kerberos authentication will fail to function on all other instances.

MS SQL Server

Overview

This is an overview of the step necessary to get your Masking Engine talking to an MS SQL Server database using Kerberos authentication. Since Active Directory already uses Kerberos for authentication, little or no additional configuration is need on the MS SQL Database server.

The following steps are described in this section:

- Create the necessary SPNs (Service Principal Names) for your MSSQL Database service in AD
- Create the DB Connector on the masking engine
- Creating a keytab for an AD User
- Troubleshooting tips

Prerequisites

Configuring cross-realm trust between Active Directory and an MIT KDC Server is a complex topic, and will not be described here. In the absence of such a setup, it is possible to make the Delphix Appliance a Kerberos client of the Active Directory (AD) Server. In this configuration, no additional KDC is necessary. The example below assumes this kind of configuration.

This section of the document uses these example values in addition to or instead of those mentioned above:

- The MSSQL server database is named `mssql-db.bar.com`.
- The AD user configured for masking access to the MSSQL database is `aduser` (rather than `krbuser` in other examples elsewhere in this document).
- The AD user that start the MS SQL Server service on the DB Server is `dbuser`.

Creating SPNs for the Database Service

MS SQL Server service will typically register several SPNs with AD upon startup. However, there are several conditions which can cause these SPNs to not be registered successfully, or to be registered with service names other than those that are expected by the Microsoft JDBC Driver for SQL Server employed by Continuous Compliance.

The service principal name for an MS SQL Server expected by Continuous Compliance is: `MSSQLSvc/`

In addition, it is **required** that a reverse mapping exist in DNS from the IP address of the MS SQL Server system to the FQDN registered.

The following commands may be run in PowerShell on the MS SQL Server to assist in debugging SPN related issues:

List all SPNs for dbuser:

```
setspn -L -U dbuser
```

Deleting an old SPN associated with dbuser:

```
setspn -U -D MSSQLSvc/other-server.ad.bar.com:SQL2008R2 dbuser
```

Here's how to create the SPN describe above:

```
setspn -U -S MSSQLSvc/mssql-db.bar.com:1433 dbuser
```

Creating the Database Connector on the Masking Engine

Once the above steps are complete, creating the database connector can be performed using the procedure above. Enter the username and optionally, password of the AD user in the Connector definition. Be sure that the AD user has sufficient access to the MS SQL Database for masking.

The password field can be left blank when creating the connector if the user is the same user configured in Server Setup for the appliance. Since keytabs are not typically used in an AD environment, it may be useful to create one manually, to avoid having a password in the DB Connector.

Creating a keytab file for an AD user

On a unix or MAC system with MIT Kerberos CLI utilities installed:

```
# ktutil

ktutil: addent -password -p krbuser -k 1 -e arcfour-hmac
<type password for krbuser>

ktutil: addent -password -p krbuser -k 1 -e aes128-cts-hmac-sha1-96
<type password for krbuser>

ktutil: addent -password -p krbuser -k 1 -e aes256-cts-hmac-sha1-96
<type password for krbuser>

ktutil: write_kt /var/tmp/krbuser.keytab

ktutil: exit

# base64 /var/tmp/krbuser.keytab ;# This is string to user for keytab in Server Setup
kerberos configuration
```



kvno doesn't matter when using Kerberos keytabs with AD. The password must match the active password for the AD user in question

Troubleshooting tips

The client uses the incorrect service name. This will typically manifest an exception mentioning cred, like:

```
Caused by: org.ietf.jgss.GSSEException: No valid credentials provided (Mechanism level: Fail to create credential. (63) - No service creds)
```

```
at sun.security.jgss.krb5.Krb5Context.initSecContext(Krb5Context.java:770)
```

```
at sun.security.jgss.GSSContextImpl.initSecContext(GSSContextImpl.java:248)
```

```
at sun.security.jgss.GSSContextImpl.initSecContext(GSSContextImpl.java:179)
```

```
at
```

```
com.microsoft.sqlserver.jdbc.KerbAuthentication.intAuthHandShake(KerbAuthentication.java:163) ... 101 common frames omitted
```

```
Caused by: sun.security.krb5.internal.KrbApErrException: Fail to create credential. (63) - No service creds at
```

```
sun.security.krb5.internal.CredentialsUtil.acquireServiceCreds(CredentialsUtil.java:162)
```

```
at sun.security.krb5.Credentials.acquireServiceCreds(Credentials.java:458)
```

```
at sun.security.jgss.krb5.Krb5Context.initSecContext(Krb5Context.java:693) ... 104 common frames omitted
```

Why might this happen: - You're using the JTDS JDBC driver, and your MSSQL Server's IP address doesn't have a reverse mapping in DNS. In this case, the driver may construct a service name like: MSSQLSvc/ : and try to use that. Either correct DNS to have a valid reverse mapping for the IP of your SQL server, or manually add an SPN to the active directory for the name the JDBC client is trying to use: - Determine the user that starts MSSQL Server on your DB machine. - From PowerShell, do: setspn -AU MSSQLSvc/ :1433 Example: setspn -AU MSSQLSvc/ 10.43.100.101:1433 AD\dbuser - The database server has multiple DNS names (FQDNs). In this case, SPNs may be registered only for some of them. It may be necessary to add SPNs for the other FQDNs as above. - The MS SQL Server didn't automatically register an SPN. There is a limit (in the thousands) to the number of SPNs that may be registered for a given AD user. It is quite possible to hit this limit in an environment where many MS SQL Server VMs are actively created and destroyed with the same configuration.

Note:

In Active Directory, setspn isn't creating a service principal with distinct key as is typical for services on MIT KDCs - rather it's mapping the service principal to the key **for** the AD user in question.

The SPN for the SQL Server is registered to the incorrect AD account

Manifests as an exception with this text: GSS failure: Defective token detected (Mechanism level: AP_REP token id does not match!)

Resolution: From PowerShell on the MS SQL Server:

```
PS> setspn -Q <SPN>
```

This will show what the user has the SPN registered.

```
PS> setspn -U -D <SPN> <WRONG_ACCT>
```

This will unregister the SPN from that user

```
PS> setspn -AU <SPN> <CORRECT_ACCT>
```

Sybase

Creating a principal and corresponding keytab on the KDC

1. SSH into the KDC as the user with sufficient privileges to run `kadmin.local`
2. Run the Kerberos configuration CLI with `kadmin.local`
3. Add a new principal you want to authenticate as later with: `add_principal <>` We're going to continue to use **krbuser** as our example Kerberos principal.
4. Once you've created the principal and provided it a password, we need to generate a keytab for it. Do so via the following command:

```
ktadd -norandkey -k v5srvtab krbuser
```

In this case, `v5srvtab` is the keytab filename, and it will be placed into whatever directory you've invoked `kadmin.local` from. Presumably, this will be the home directory of the machine.

1. You now have everything you need done on the KDC, but you will need your keytab file later as well as the **krb5.conf** file that is located in the home directory of the KDC, so consider moving them somewhere (probably your local machine) that will be convenient for you to access later.

Configuring the Sybase image for Kerberos

1. Startup a Sybase database.
2. **Note:** Each Sybase database machine may have multiple Sybase instances running on it at a given point in time. In this case, I am configuring the `ASE_1550_S5` instance, but these steps can be done on any instance so long as you change the `$SYBASE_HOME` directories accordingly.
3. Connect to the particular Sybase instance you are working on and invoke the following sql statement:
`sp_configure 'use security services', 1`
4. Continue to create a user with the same name as the principal name you created previously on the KDC, in this case **krbuser**: `sp_addlogin krbuser, <password>`
5. Change your **\$SYBASE** environment variable to point to the Sybase directory for whichever instance you are configuring. In this case, we want to do: `export SYBASE=/opt/sybase/15-5`
6. Open the **\$SYBASE/interfaces file**, and find the header for whichever Sybase instance you are configuring. In our case, it is **ASE_1550_S5**. You should see something that looks like this: `ASE1550_S5`

```
`master tcp ether 10.43.89.241 5500`
`master tcp ether localhost 5500`
`query tcp ether 10.43.89.241 5500`
`query tcp ether localhost 5500`
```

You want to add the following line to this:

```
secmech 1.3.6.1.4.1.897.4.6.6
```

This line is **static**, while the other lines in **this** section are dynamically generated **for** your instance. So, your **final** result should look something like **this**:

```
ASE1550_S5
```

```
master tcp ether 10.43.89.241 5500 < your numbers will vary
```

```
master tcp ether localhost 5500 < your numbers will vary
```

```
query tcp ether 10.43.89.241 5500 < your numbers will vary
```

```
query tcp ether localhost 5500 < your numbers will vary
```

1. Navigate to **\$SYBASE/OCS-15_0/config**. You should see **libtcl64.cfg** and **libtcl.cfg**
2. Change the contents of **libtcl64.cfg** to be this:

```
`[DIRECTORY]`
`;ldap=libsybdldap.so ldap://ldaphost/dc=sybase,dc=com`
`[SECURITY]`
`csfkrb5=libsybskrb64.so secbase=@bar.com libgss=/lib64/libgssapi_krb5.so.2.2`
```

```
[FILTERS];ssl=libsybfssl.so`
```

2. Change the contents of **libtcl.cfg** to be this:

```
`[DIRECTORY]`
`;ldap=libsybdldap.so ldap://ldaphost/dc=sybase,dc=com`
`[SECURITY]`
`csfkrb5=libsybskrb.so secbase=@bar.com libgss=/lib64/libgssapi_krb5.so.2.2`
`[FILTERS]`
`;ssl=libsybfssl.so`
```

3. **Note** that the **@bar.com** value is our realm name that is determined by the KDC. Realistically, you should never have to deal with this, and it should never change, but if for some reason it does, that value needs to be updated.

1. Create a directory for those Kerberos config files you created on the KDC in the previous set of steps:

```
sudo mkdir /krb
```

Copy into **/krb** your keytab file **v5srvtab** and config file **krb5.conf** that you took off of the KDC earlier.

1. Head to **\$SYBASE/ASE-15_0/install** and open the **RUN_ASE1550_S5** file. We're going to add information so that Sybase knows where to find our keytab and our **krb5.conf** file, so change the content to look like this:

```
#!/bin/sh

#

# ASE page size (KB) : 4096

# Master device path: /opt/sybase/devices/data5/S5_master.dat

# Error log path: /opt/sybase/errorlogs/ASE1550_S5.log

# Configuration file path: /opt/sybase/15-5/ASE-15_0/ASE1550_S5.cfg

# Directory for shared memory files: /opt/sybase/15-5/ASE-15_0
```

```
# Adaptive Server name: ASE1550_S5

#

export **KRB5_KTNAME**=/krb/v5srvtab

export **KRB5_CONFIG**=/krb/krb5.conf

/opt/sybase/15-5/ASE-15_0/bin/dataserver \

-kASE1550_S5@bar.com \

-d/opt/sybase/devices/data5/S5_master.dat \

-e/opt/sybase/errorlogs/ASE1550_S5.log \

-c/opt/sybase/15-5/ASE-15_0/ASE1550_S5.cfg \

-M/opt/sybase/15-5/ASE-15_0 \

-sASE1550_S5 \
```

1. Reboot the Sybase instance you're working so that it reads in all of these configuration changes.
2. Connect to the Sybase instance as the **dbo** user so that you may give dbo privileges to your Kerberos authentication login on a particular database within the instance. Below is an example of doing so with the database **potatoes**:

```
>> sql5

1> use potatoes

2> go

1> sp_addalias instructions, dbo

2> go

Alias user added.

(return status = 0)
```

1. Now, to access the Sybase instance via Kerberos and confirm success, you can do the following set of commands (I put these three lines into a script called **connect.sh** for future convenience):

```
#!/bin/sh  
  
kinit -k -t /krb/v5srvtab <>  
  
export SYBASE='/opt/sybase/15-5'  
  
/opt/sybase/15-5/OCS-15_0/bin/isql64 -V -SASE1550_S5
```

Testing by creating a Kerberos connector on the Delphix Engine

1. Start by configuring your engine for Kerberos. SSH into the engine as the Delphix user and run the following command: `/opt/delphix/server/bin/jmxtool tunable set enabled_features KERBEROS true`
2. Log into the Delphix Engine and proceed through the first-time setup.
3. Once the first-time setup is complete, log into the Delphix Setup page, proceed to Preferences > Kerberos Configuration. Add the information for your KDC to configure it with the principal name you created earlier, **krbuser**. You can get the keytab by running the following command on your keytab file: `base64 v5srvtab`

Copy the output as plaintext into the keytab field of the Kerberos configuration box.

Finally, create a Sybase connector with parameters that look like this, and if your “test connection” attempt succeeds you’re all set!

Create Connection

Type

Database - Sybase

Basic Advanced

Connection Name

Sybase Kerberos

Port

4000

Schema Name

dbos

Use Kerberos Authentication

Principal Name

krbuser

Database Name

potatoes

Service Principal

ASE1550_S5

Host Name/ IP

sybaseHostName.bar.com

Password

LEAVE BLANK TO USE KEYTAB

Custom Properties File

Select...

Test Connection

Cancel

Save

Password vault configuration

Introduction

As of release 10.0.0.0, the Continuous Compliance Engine supports the use of HashiCorp and CyberArk password vaults for connections to PostgreSQL, MSSQL and Oracle databases. Utilizing this feature requires the presence of either a HashiCorp or CyberArk vault, as well as additional configuration actions on the Continuous Compliance Engine.

 Password vault authentication is not supported for containerized masking deployments at this time.

Configuring a password vault on the appliance

Before attempting to access a password vault, the CA certificate for the vault must first be added to the Compliance Engine's trust store. Certificates can be managed through the Delphix Server Setup UI and the steps for doing so can be found [here](#).

Currently, password vaults and the associated credential paths can only be configured on the appliance using the API. The Continuous Compliance Engine's web API includes two endpoints, `password-vaults` and `credential-paths` for managing the setup of vaults and credentials.

Setting up a password vault

The POST action on the `password-vaults` endpoint is used to provide information on the type of vault to be accessed and the location of the server hosting the vault.

For a HashiCorp vault, the body of the request will be similar to:

```
{
  "name": "HashiVault",
  "vaultType": "HASHICORP",
  "configJson": {
    "host": "123.45.67.89",
    "port": 8200,
    "namespace": "sample/child",
    "authType": "TOKEN",
    "token": "hvs.kvITvwsI4gs"
  },
  "description": "Vault description is optional"
}
```

 Namespaces are only relevant when using the Enterprise version of the HashiCorp product. If this field is specified, it should match the namespace being used on the HashiCorp server.

To use either AppRole or Certificate based authentication, the following substitutions can be made to the above example:

```
"authType": "APPROLE",
  "roleId": "your-role-id",
  "secretId": "your-secret"
```

or

```
"authType": "CERTIFICATE",
  "certificate": "-----BEGIN CERTIFICATE-----\nMIa1ZqA=\n-----END CERTIFICATE-----",
  "privateKey": "-----BEGIN RSA PRIVATE KEY-----\nUw9aPq\n-----END RSA PRIVATE
KEY-----",
  "roleName": "sampleRole"
```

For CyberArk, the request body will be similar to:

```
{
  "name": "CyberVault",
  "vaultType": "CYBERARK",
  "configJson": {
    "host": "cyberark01.myserver.com",
    "port": 443,
    "appId": "MyApp",
    "authType": "CERTIFICATE",
    "certificate": "-----BEGIN CERTIFICATE-----\nMIa1ZqA=\n-----END CERTIFICATE-----"
    "privateKey": "-----BEGIN PRIVATE KEY-----\nMIa1ZqA=\n-----END PRIVATE KEY-----"
  },
  "description": "Vault description is optional"
}
```

Setting up a credential path

Credential paths are used to specify the location of the credentials within a password vault.

HashiCorp

The Continuous Compliance Engine currently supports two types of HashiCorp secrets engines: *database* and *key-value-v2*.

The request body for a HashiCorp credential path will be similar to:

```
{
  "credentialPathName": "HashiCredentialPath",
  "description": "Credential path description is optional",
  "passwordVaultId": 1,
  "credentialParameters": {
    "engineType": "KEY_VALUE_V2",
    "engine": "secret-engine-name",
    "path": "secret-path",
    "usernameKey": "username",
    "passwordKey": "password"
  }
}
```

```
}

```

Database secrets engines support dynamic secrets by generating database credentials based on configured roles. When using a database secrets engine, set "engineType" to "DATABASE" and use "role" to specify the name of the role to create credentials against.

```
"credentialParameters": {
  "engineType": "DATABASE",
  "engine": "database-engine-name",
  "role": "my-role",
  "usernameKey": "username",
  "passwordKey": "password"
}
```

CyberArk

The request body for a CyberArk credential path will be similar to:

```
{
  "credentialPathName": "CyberCredentialPath",
  "description": "Credential path description is optional",
  "passwordVaultId": 1,
  "credentialParameters": {
    "queryString": "Safe=DevTest;Folder=Root;Object=postgres01"
  }
}
```

Configuring the database connector

Database connectors can be configured to use a password vault through either the Continuous Compliance Engine UI or the APIs.

UI configuration

When creating or editing a PostgreSQL, MSSQL or Oracle database connector, check the **Use Password Vault** option and then select the required credential path from the **Credential Path** dropdown. If the 'Test Connection' run succeeds then it is complete.

API configuration

`CredentialPathId` is an optional field when creating a PostgreSQL, MSSQL or Oracle Database Connector via the API. Setting this value to the id of an existing credential path object will result in the connector using password vaults to retrieve the credential. As an example:

```
{
  "connectorName": "psql-connector",
  "databaseType": "POSTGRES",
  "environmentId": 1,
  "host": "mpv-psql.mydb.co",

```

```
"port": 5432,  
"databaseName": "postgres",  
"schemaName": "public",  
"credentialPathId": 1  
}
```

DB2 connector license installation

[-] Use of IBM's custom driver is disabled in containerized masking. Access to Linux DB2 databases is still possible with the out-of-the-box drivers provided.

If you have been licensed to use the Continuous Compliance DB2 Connector for Mainframe or DB2 Connector for iSeries, you will need to obtain the respective DB2 Connector package (tar file) and apply it to your Masking Engine(s). Each package is intended to be installed and run from a workstation or laptop, not from the Delphix Appliance. These packages contain a script that must be used in a bash shell and depends on the availability of the **curl** and **ssh** commands to install the respective license on your remote Delphix Appliance.

Applying DB2 connector for mainframe

1. Go to the <https://download.delphix.com/folder/580/Delphix%20Product%20Releases/DB2%20Masking%20Mainframe> and download `DB2MaskingMainframe.tar`
2. Extract its contents using `tar -xvf DB2MaskingMainframe.tar`
3. `cd db2-license`
4. `./installdb2license.sh -h MASKING_ENGINE_HOST -P MASKING_ENGINE_PORT -u MASKING_ENGINE_ADMIN_USERNAME -p MASKING_ENGINE_ADMIN_PASSWORD [-C MASKING_ENGINE_PUBLIC_KEY_FILE]`

Where:

MASKING_ENGINE_HOST is the hostname for where the masking engine is running.

MASKING_ENGINE_PORT is the port for where the masking engine is listening on the **MASKING_ENGINE_HOST** (default is port 80).

MASKING_ENGINE_ADMIN_USERNAME is the username for connecting to the masking engine (e.g., `delphix_admin`).

MASKING_ENGINE_ADMIN_PASSWORD is the masking engine password for

MASKING_ENGINE_PUBLIC_KEY_FILE is the optional trusted server certificate (server public key) obtained from the masking engine.

[-] To run the enablement script securely, run `installdb2license.sh` specifying your secure port (e.g., 8443) and trusted server certificate (server public key) using the `-C` option.

The script will enable the DB2 Mainframe connector and then recycle the Masking Engine, prompting the user for the Delphix sysadmin password for to first stop the Masking Engine and then to start it. After the `DB2MaskingMainframe.tar` package has been applied to your Masking Engine(s), "Database - MAINFRAME DB2" will appear in the Connector drop-down of the Masking Engine UI and can be used in the same way as other Database Connectors to create, profile, mask, certify, and provision rulesets.

Applying DB2 connector for iSeries

1. Go to the [Delphix Download site](#) and download `DB2MaskingISeries.tar`
2. Extract its contents using `tar -xvf DB2MaskingISeries.tar`
3. `cd db2-license`

4. `./installdb2license.sh -h MASKING_ENGINE_HOST -P MASKING_ENGINE_PORT -u MASKING_ENGINE_ADMIN_USERNAME -p MASKING_ENGINE_ADMIN_PASSWORD [-C MASKING_ENGINE_PUBLIC_KEY_FILE]`

Where:

MASKING_ENGINE_HOST is the hostname for where the masking engine is running.

MASKING_ENGINE_PORT is the port for where the masking engine is listening on the MASKING_ENGINE_HOST (default is port 80).

MASKING_ENGINE_ADMIN_USERNAME is the username for connecting to the masking engine (default is delphix_admin).

MASKING_ENGINE_ADMIN_PASSWORD is the masking engine password for MASKING_ENGINE_ADMIN_USERNAME.

MASKING_ENGINE_PUBLIC_KEY_FILE is the optional trusted server certificate (server public key) obtained from the masking engine.



To run the enablement script securely, run `installdb2license.sh` specifying your secure port (e.g., 8443) and trusted server certificate (server public key) using the `-C` option.

The script will enable the DB2 iSeries connector and then recycle the Masking Engine, prompting you for the Delphix sysadmin password to first stop the Masking Engine and then start it. After the DB2MaskingISeries.tar package has been applied to your Masking Engine(s), "Database - ISeries DB2" will appear in the Connector drop-down of the Masking Engine UI and can be used in the same way as other Database Connectors to create, a profile, mask, certify, and provision rulesets.

Continuous Compliance Engine icon reference

This topic illustrates the icons that appear on the Continuous Compliance Engine graphic user interface and describes the meaning of each.

Icon	Description
	Edit
	Export
	Copy
	Delete
	Job Success
	Job Created
	Mask
	Run Job
	Ruleset Refresh
	Ruleset refresh not applicable for file rulesets
	Job Running
	Cancel Job
	Ruleset Refresh in Progress

Delphix masking terminology

Before getting started with the Continuous Compliance Engine, an overview of universal terms and concepts will build and unify how different masking components come together. The following provides a brief overview of the key concepts within the masking service.

High level concepts

These concepts are the high level concepts users run into.

Term	Definition
Application	An Application is a tag that is assigned to one or more environments. We recommend using an application name that is the same as the application associated with the environments.
Connector	Connectors are any set of data (database, file, etc) that have been connected to the Delphix Data Platform. These data sources can be physical or virtualized data sources.
Domain	A domain represents a correlation between various sensitive data categories (social security numbers) and the way it should be secured.
Environment	An environment is a construct that can be used to describe a collection of masking jobs associated with a group of data sources.
In-place	In-place masking is 1 of 2 procedures that can be used to apply masking algorithms to a data source. By choosing the In-place option, Delphix will read data from the data source, secure the data in the Engine and then update the data source with the secure data.
On-the-fly	On-the-fly masking is the second procedure that can be used to apply masking algorithms to a data source. By choosing the On-the-fly option, Delphix will read data from the data source, secure the data in the Engine and then place the secure data in a target source (different from the location of the original data source).
Inventory	An inventory describes all of the data present in a particular data source and defines the methods which will be used to secure it. Inventories typically include the table name, column name, the data classification, and the chosen algorithm.
Profile	Profiling uses a variety of different methods to classify data in a data source into different categories. These categories are known as domains. The profile process also assigns recommended algorithms for securing the data based on the the domain.
Ruleset	A rule set is group of tables or flat files within a particular data source that a user may choose to run profile, masking, or tokenization jobs on.

Masking algorithms

The following terminology is around the different Algorithms that users may use to secure their data.

Term	Definition
Algorithm Framework	A type of masking algorithm. One or more usable instances of an <i>algorithm framework</i> may be created. For example, "FIRST NAME SL" is an instance of the Secure Lookup <i>algorithm framework</i> .
Algorithm Instance	A named combination of algorithm framework and configuration values. <i>Algorithm instances</i> are applied to data fields and columns in the inventory in order to mask data.
Built-in Algorithm	An algorithm instance or framework included with the Masking Engine software. This includes several built-in algorithm instances that provide masking behavior that doesn't correspond to any built-in algorithm framework.
Non-conformant Data	Some masking algorithms require data to be in a particular format. The required format may vary by the configuration of the algorithm instance. For example, a particular Segment Mapping algorithm might be configured to expect a 10 digit number. Data which doesn't fit the pattern expected by an algorithm is called <i>nonconforming data</i> or <i>non-conformant data</i> . By default, <i>non-conformant data</i> is not masked, and warnings are recorded for the masking job. Warnings are indicated by a yellow triangle warning marker next to the job execution in Environment and Job Monitor pages. Whether <i>non-conformant data</i> results in a warning or failure is configurable for each algorithm instance.
Collision	The term <i>collision</i> describes the case where a masking algorithm masks two or more unique input values to the same output value. For example, a first name Secure Lookup algorithm might mask both "Amy" and "Jane" to the same masked value "Beth". This may be desirable, in the sense that it further obfuscates the original data, however <i>collisions</i> are problematic for data columns with uniqueness constraints.
Secure Lookup	The most commonly used algorithm framework. Secure lookup works by replacing each data value with a new value chosen from an input file. Replacement values are chosen based on a cryptographic hash of the original value, so masking output is consistent for each input. Secure lookup algorithms are easy to configure and work with different languages. When this algorithm replaces real data with fictional data, <i>collisions</i> , described above, are possible. Because many types of data, such as first or last name, address, etc, are not unique in real data, this is often acceptable. However, if unique masking output for each unique input is required, consider using a mapping or segment mapping algorithm, described below.
Segment Mapping	This algorithm permutes short numeric or alpha-numeric values to other values of the same format. This algorithm is guaranteed to not produce collisions, so long as the set of permissible mask values is at least as large as input or "real" set. The maximum number of digits or characters in the masked value is 36. You might use this method if you need columns with unique values, such as Social Security Numbers, primary key columns, or foreign key columns.

Term	Definition
Mapping	<p>Similar to secure lookup, a mapping algorithm allows you to provide a set of values that will replace the original data. There will be no collisions in the masked data, because each input is always matched to the same output, and each output value is only assigned to one input value. In order to accomplish this, the algorithm records, in an encrypted format, all known input to output mappings.</p> <p>You can use a mapping algorithm on any set of values, of any length, but you must know how many values you plan to mask, and provide a set of unique replacement values sufficient to replace each unique input value.</p> <p>NOTE: When you use a mapping algorithm, you cannot mask more than one table at a time. You must mask tables serially.</p>
Binary Lookup	<p>Replaces objects that appear in object columns. For example, if a bank has an object column that stores images of checks, you can use binary lookup algorithm to mask those images. The Delphix Engine cannot change data within images themselves, such as the name on X-rays or driver's licenses. However, you can replace all such images with a new, fictional image. This fictional image is provided by the owner of the original data.</p>
Tokenization	<p>The only type of algorithm that allows you to reverse its masking. For example, you can use a tokenization algorithm to mask data before you send it to an external vendor for analysis. The vendor can then identify accounts that need attention without having any access to the original, sensitive data. Once you have the vendor's feedback, you can reverse the masking and take action on the appropriate accounts.</p> <p>Like mapping, a tokenization algorithm creates a unique token for each input such as "David" or "Melissa." The Delphix Engine stores both the token and original so that you can reverse masking later.</p>
Min Max	<p>Values that are extremely high or low in certain categories allow viewers to infer someone's identity, even if their name has been masked. For example, a salary of \$1 suggests a company's CEO, and some age ranges suggest higher insurance risk. You can use a min max algorithm to move all values of this kind into the midrange.</p>
Data Cleaning	<p>Does not perform any masking. Instead, it standardizes varied spellings, misspellings, and abbreviation for the same name. For example, "Ariz," "Az," and "Arizona" can all be cleaned to "AZ."</p>

Term	Definition
Free Text Redaction	<p>Helps you remove sensitive data that appears in free-text columns such as “Notes.” This type of algorithm requires some expertise to use, because you must set it to recognize sensitive data within a block of text.</p> <p>One challenge is that individual words might not be sensitive on their own, but together they may be. This algorithm uses profiler sets to determine which information it needs to mask. You can decide which expressions the algorithm uses to search for material such as addresses. For example, you can set the algorithm to look for “St,” “Cir,” “Blvd,” and other words that suggest an address. You can also use pattern matching to identify potential sensitive information. For example, a number that takes the form 123-45-6789 is likely to be a Social Security Number.</p> <p>You can use free text redaction algorithm to show or hide information by displaying either a “deny list” or an “allow list.”</p>

Profile job concepts

The following set of concepts are options available to the user for configuring a profiling job.

Term	Definition
Job Name	A free-form name for the job you are creating. Must be unique.
Multi-Tenant	Check the box if the job is for a multi-tenant database. This option allows existing rulesets to be reused to mask identical schemas via different connectors. The connector can be selected at job execution time.
Rule Set	Select a ruleset that this job will execute against.
No. of Streams	The number of parallel streams to use when running the jobs. For example, you can select two streams to run two tables in the ruleset concurrently in the job instead of one table at a time.
Min Memory (MB) <i>optional</i>	Minimum amount of memory to allocate for the job, in megabytes.
Max Memory (MB) <i>optional</i>	Maximum amount of memory to allocate for the job, in megabytes.

Term	Definition
Feedback Size <i>optional</i>	The number of rows to process before writing a message to the log. Set this parameter to the appropriate level of detail required for monitoring your job. For example, if you set this number significantly higher than the actual number of rows in a job, the progress for that job will only show 0 or 100%
Profile Sets <i>optional</i>	The name of a profile set, which is a subset of expressions (for example, a subset of financial expressions).
Comments <i>optional</i>	Add comments related to this job.
Email <i>optional</i>	Add email address(es) to which to send status messages. Separate addresses with a comma (,).

Masking job concepts

These concepts are options available to the user for configuring a masking job.

Term	Definition
Job Name	A free-form name for the job you are creating. Must be unique across the entire application.
Masking Method	Select either In-Place or On-The-Fly.
Multi-Tenant	Check the box if the job is for a multi-tenant database. This option allows existing rulesets to be reused to mask identical schemas via different connectors. The connector can be selected at job execution time.
Rule Set	Select a ruleset for this job to execute against.
Masking Method	Select either In-place or On-the-fly.
Min Memory (MB) optional	Minimum amount of memory to allocate for the job, in megabytes.
Max Memory (MB) optional	Maximum amount of memory to allocate for the job, in megabytes.

Term	Definition
Update Threads	The number of update threads to run in parallel to update the target database. For database using T-SQL, multiple update/insert threads can cause deadlock. If you see this type of error, reduce the number of threads that you specify in this box.
Commit Size	The number of rows to process before issuing a commit to the database.
Feedback Size	The number of rows to process before writing a message to the logs. Set this parameter to the appropriate level of detail required for monitoring your job. For example, if you set this number significantly higher than the actual number of rows in a job, the progress that job will show 0% or 100%.
Disable Trigger <i>optional</i>	Whether to automatically disable database triggers. The default is for this check box to be clear and therefore not perform automatic disabling of triggers.
Drop Index <i>optional</i>	Whether to automatically drop indexes on columns which are being masked and automatically re-create the index when the masking job is completed. The default is for this check box to be clear and therefore not perform automatic dropping of indexes.
Prescript <i>optional</i>	Specify the full pathname of a file that contains SQL statements to run before the job starts, or click Browse to specify a file. If you are editing the job and a pre script file is already specified, you can click the Delete button to remove the file. (The Delete button only appears if a prescript file was already specified.)
Postscript <i>optional</i>	Specify the full pathname of a file that contains SQL statements to be run after the job finishes, or click Browse to specify a file. If you are editing the job and a postscript file is already specified, you can click the Delete button to remove the file. (The Delete button only appears if a postscript file was already specified.)
Comments <i>optional</i>	Add comments related to this masking job.
Email <i>optional</i>	Add email address(es) to which to send status messages.

Changing the IP address of the Delphix Engine

You can change the IP address of the Delphix Engine either from the User Interface or using the Command-Line Interface.

- For Containerized Masking, networking is handled via the underlying kubernetes infrastructure. There is no interface through the application to change the IP address. Changing the IP address of a containerized instance requires those changes to happen in kubernetes and its underlying nameservice. Frequently it is managed by the network proxy that directs traffic to the containerized instance.

Pre-requisites

- Ensure that no masking jobs are running.

Changing the IP address from the user interface

Perform the following procedure to change the IP address of the Delphix Engine from the UI.

1. Launch the Delphix Setup application.
2. Go to **System > Server Setup** in the Delphix Management interface, or click **Server Setup** in the Delphix Engine login screen.
3. In the **Network** panel, click **Modify**.
4. Under **DNS Services**, enter the new IP address.
5. Click **Ok**.
6. Refresh all environments by clicking the **Refresh** option on the Environments screen.

Changing the IP address using CLI

Perform the following procedure to change the IP address of the Delphix Engine using CLI.

1. Log into the Delphix CLI using your sysadmin account.

```
delphix> network
delphix network> setup
delphix network interface> list
NAME
vmxnet3s0
delphix network interface> select vmxnet3s0
delphix network interface 'vmxnet3s0'> get
  type: NetworkInterface
  name: vmxnet3s0
  addresses:
    0:
      type: InterfaceAddress
      address: 10.1.2.3/24
      addressType: STATIC
      enableSSH: true
      state: OK
  dataNode: DATA_NODE-34
  device: vmxnet3s0
```

```

macAddress: 0:c:29:32:96:a3
mtu: 1500
mtuRange: 60-9000
reference: NETWORK_INTERFACE-vmxnet3s0-DATA_NODE-34
state: OK

```

2. Run the update command and update the address to the new IP address for the Delphix Engine.

```

delphix network interface 'vmxnet3s0'> update
delphix network interface 'vmxnet3s0' update *> edit addresses.0
delphix network interface 'vmxnet3s0' update addresses.0 *> get
Properties
  type: InterfaceAddress
  address: 172.16.151.154/24
  addressType: STATIC
  enableSSH: true

delphix network interface 'vmxnet3s0' update addresses.0 *> set address=10.1.2.4/24
delphix network interface 'vmxnet3s0' update addresses.0 *> get
  type: InterfaceAddress (*)
  address: 10.1.2.4/24 (*)
  addressType: STATIC (*)
  enableSSH: true (*)

```

3. Commit the operation.

```

delphix network interface 'vmxnet3s0' update addresses.0 *> commit
delphix network interface 'vmxnet3s0'> get
  type: NetworkInterface
  name: vmxnet3s0
  addresses:
    0:
      type: InterfaceAddress
      address: 10.1.2.4/24
      addressType: STATIC
      enableSSH: true
      state: OK
  dataNode: DATA_NODE-34
  device: vmxnet3s0
  macAddress: 0:c:29:32:96:a3
  mtu: 1500
  mtuRange: 60-9000
  reference: NETWORK_INTERFACE-vmxnet3s0-DATA_NODE-34
  state: OK

```

Stopping and starting the containerized Continuous Compliance Engine

Overview

This article describes how to stop and start the containerized Delphix Continuous Compliance engine. For information on performing the tasks for the Virtual Machine Masking Engine, please see the documentation located in the document [Starting, Stopping, and Restarting the Masking Engine](#).

Containerized deployments are dependent on a customer-created configuration file which can be named anything. For the purposes of this document, the default name of `kubernetes-config.yaml` will be used. Also, any command-line examples will assume that this file is in the current directory to simplify the example.

Starting the containerized Masking Engine

Starting the engine is a simple matter of asking Kubernetes to create the Pod described by the Pod configuration file. This is done with a single `kubectl` command.

```
$ kubectl create -f ./kubernetes-config.yaml
```

The Pod will take some time to start. The status of the Pod can be verified with another simple `kubectl` command.

```
$ kubectl get pods
NAME                READY   STATUS    RESTARTS   AGE
delphix-masking-0   3/3     Running   2 (5d14h ago)  13d
```

A Containerized Masking Pod consists of 3 containers. The above output demonstrates a Pod where 3/3 containers are `READY`. This is a Pod that is up and running and ready to accept connections.

 It is common for the first 2 containers of the Pod to enter a `READY` state very quickly and for the 3rd container to take some time to become ready. How long is dependent on a number of factors including the underlying infrastructure. (how powerful, how busy)

If the Pod `STATUS` indicates an error or the number of restarts is consistently climbing, that indicates that there is a problem with the Pod and debugging will need to be done to determine the problem and the appropriate resolution.

Stopping the containerized Masking Engine

Stopping a running Pod is simple despite some confusing terminology. The Kubernetes terminology for stopping a Pod is `delete`, but this command does not delete any of the containers or persistent volumes. It only stops the running Pod. The command to stop a running pod is of the same form as starting the Pod.

```
$ kubectl delete -f ./kubernetes-config.yaml
```

This tells Kubernetes to shut down whatever it previously started. Because the Containerized Masking Engine is a Stateful application, it has persistent storage. This persistent storage is not deleted when the Pod is shut down.

If a Pod that was shut down is then restarted, it will attempt to re-attach any persistent storage defined in the `kubernetes-config.yaml` file.

Removing persistent volumes / persistent volume claims

If it is necessary to delete any persistent volumes (PVs) and persistent volume claims (PVCs) associated with the Pod, that will have to be done manually. It is possible to locate any PVs and PVCs that exist with some simple `kubectl` commands.

```
$ kubectl get pv
NAME                                CAPACITY  ACCESS MODES  RECLAIM POLICY  STORAGECLASS
STATUS  CLAIM
REASON  AGE
pvc-7fe1d352-de17-4132-b2a1-152f4e9cfefc  20Gi      RWX           Delete          microk8s-hostpath
Bound   container-registry/registry-claim
183d
pvc-a7275ce3-b630-4d4c-9712-b5124358cb7f  4Gi       RWO           Delete          microk8s-hostpath
Bound   default/masking-persistent-storage-delphix-masking-0
15d
nfs-pv                                500Mi     RWO           Retain          nfs-storage
Bound   default/nfs-pvc
13d

$ kubectl get pvc
NAME                                STATUS  VOLUME
CAPACITY  ACCESS MODES  STORAGECLASS  AGE
masking-persistent-storage-delphix-masking-0  Bound   pvc-a7275ce3-b630-4d4c-9712-
b5124358cb7f  4Gi          RWO           microk8s-hostpath  15d
nfs-pvc                                Bound   nfs-pv
500Mi       RWO          nfs-storage   13d
```

To completely remove a Pod (a clean slate) would require the removal of any PVs and PVCs associated with the Pod. The first step is to shut down the Pod. Once the Pod is no longer running, removing the PVC will frequently also remove the associated PV.

Removing either the PV or PVC is a simple matter of using the appropriate `kubectl` command. To illustrate removing a PVC, simply take note of the name of the object. From the example output above, there is a PVC named `masking-persistent-storage-delphix-masking-0`. To remove it, use the following command.

```
$ kubectl delete pvc masking-persistent-storage-delphix-masking-0
```

Stopping, starting, and restarting the continuous compliance engine

Overview

This article describes how to stop, start, and restart the Virtual Machine-based Delphix Continuous Compliance engine. Use cases, troubleshooting tips before a restart, and steps in the CLI are outlined in the following sections. For instructions on stopping and starting the Containerized Masking Engine, please see the document [Stopping and Starting the Containerized Masking Engine](#).

Use cases examples

Stopping and starting the Masking Engine may be required when performing:

- Masking Engine maintenance work.
- Backup and Restore.

Restarting the Masking Engine may be required if:

- The Masking Engine is unreachable or unresponsive.
- A Masking Job is in an incorrect state.

 Stopping and Starting the Masking Engine will terminate all running jobs; this includes Imports, Inventory Scans, Profiling and Masking Jobs, etc.

Troubleshooting before a restart

If the Masking Engine is unreachable, the following should always be checked before a restart:

- Verify that the Engine is reachable over the network using ping.

 Verify that no jobs are running (unless the job should be terminated). If a root cause investigation is needed, please open a case with Delphix Support and upload a support bundle.

 Containerized Masking functions very differently from the Virtual Machine deployment. For information on performing these same functions for Containerized Masking, please see the documentation page for [Stopping and Starting the Containerized Masking Engine](#).

Using the shell or putty, access the Masking Engine and login using the sysadmin user.

 The sysadmin password is the password set when the Masking Engine was configured.

```
# Access CLI using SSH.  
ssh sysadmin@<yourEngine>
```

Using the Command-Line Interface (CLI)

The CLI provides means to access information and execute commands on the Engine without a GUI; one of which is to stop and start the Continuous Compliance Engine. This is done using the system menu.

1. At the CLI prompt, type **system**.
2. At the system prompt, do one of the following, depending on the desired action:
 - a. To enable the engine: type **startMasking** and then commit.
 - b. To disable the engine: type **stopMasking** and then commit.
 - c. To restart the engine: type **stopMasking** and commit, then **startMasking** and commit.
3. To exit the CLI, type **exit**.

 If the Masking Engine fails to start, it could be worth waiting a few minutes (2 minutes or so) and then try `stopMasking`, followed by `startMasking` again. Startup failure could be the masking service entering Maintenance Mode. You cannot clear Maintenance Mode by entering `startMasking`; you must use `stopMasking`, followed by `startMasking`. If this fails, Delphix Support needs to investigate why the service failed.

Restarting the Masking Engine example

Below is an example of how to restart the Continuous Compliance Engine using the CLI.

```
$ ssh sysadmin@yourEngine
Password:
yourEngine> system
yourEngine system> ls
startMasking stopMasking
yourEngine system> stopMasking
yourEngine system stopMasking *> commit
yourEngine system> startMasking
yourEngine system startMasking *> commit
yourEngine system> exit
Connection to yourEngine closed
```

Upgrading the Delphix Continuous Compliance Engine

Upgrades for virtual machine engines

Upgrading Delphix Engine appliances is a multi-step process. This process will affect the availability of the Delphix Engine administrative interface and virtual datasets during the operation, based on the type of upgrade chosen.

 Customers running version 5.3.9 and earlier that are requesting an upgrade to 6.0.0.0 and above, please contact Delphix Support to help coordinate this upgrade. Upgrading from 6.0.x to 6.0.x includes pre-checks packaged in the upgrade image, thus, contacting Delphix Support is **not required** for this upgrade (e.g. 6.0.0.0 -> 6.0.9.0).

For more information on upgrades and the process, please visit the [Upgrade section](#) of the Virtualization documentation.

Upgrades for containerized Masking Engines

Containerized Masking is generally expected to be used in an ephemeral fashion. The general process for utilizing newer versions is to upload the new set of containers and deploy new engines from them.

There is not currently a certified process by which to upgrade a Containerized Masking engine in-place. If you believe that you have a need for such, please contact your Delphix Representative and request that they open an enhancement request.

Utilization

Overview

Delphix helps with keeping track of Compliance engine utilization. You can access the utilization reports from the **Admin > Utilization** page. This article discusses the UI.

Utilization UI page

The **Utilization** page can be found in the UI under the **Admin** tab. This page provides the user the capability to generate two types of utilization reports in PDF format:

- **Jobs:** The report lists the number of job executions for a specified environment and date range.
- **Database Size:** This report is designed to assist customers on usage based pricing contracts. The report contains the databases masked and their respective size.
 - Currently, this information is only gathered for Oracle databases (Oracle 12c or greater). Support for Oracle 11g and other database platforms will be added in future releases.
 - There is also an API endpoint to get database usage information. For more information, please refer to [API documentation](#).

The jobs utilization report

To generate the jobs utilization report:

1. Select the **Utilization Type** as Jobs.
2. Select the **Job Environment** from the list.
3. Provide **Start Date** and **End Date**.

Home > Admin > Utilization

Admin

Utilization

The screenshot shows the Utilization page in the Admin section. On the left is a navigation menu with options: Users, Email Notification, Utilization (highlighted), Logs, and About. The main content area contains three input fields: 'Utilization Type' set to 'Jobs', 'Select Job Environment' set to 'My Staging Environment', and 'Start Date - End Date' set to '04/01/2023 - 04/30/2023'. A blue 'Generate PDF' button is located below these fields.

4. Click **Generate PDF**.

The database size report

To generate the database size report:

1. Select the **Utilization Type** as Database Size.

2. Select a Date Range from the predefined list of date ranges. Select Custom to supply a custom Start Date and End Date.

Home > Admin > Utilization

Admin

Utilization

[Users](#)

[Email Notification](#)

[Utilization](#)

[Logs](#)

[About](#)

Utilization Type

Database Size x ▾

Date Ranges

2023 Apr - Jun (Q2) x ▾

Start Date - End Date

04/01/2023 – 06/30/2023 📅

[Generate PDF](#)

3. Click **Generate PDF**.

Support matrix

The Delphix engine supports database size calculation for the following databases:

Oracle

Data Source	Version	Availability	Setup Instructions
Oracle	Oracle 11g	Unavailable	
	Oracle 12c and later	Available	Preparing Oracle Database for Profiling/Masking
All other data sources		Unavailable	

Multiple options can be selected.

Preparing data

This section includes the following topics:

- [Database user permissions for executing masking and profiling job](#)
- [Preparing Oracle database for profiling/masking](#)
- [Preparing SQL server database for profiling and masking](#)
- [Preparing Sybase database for profiling and maskin](#)

Database user permissions for executing masking and profiling job

Introduction

This section provides the recommended list of permissions required for executing Masking and Profiling jobs on the Continuous Compliance Engine. This page provides general permission recommendations. The subsequent pages in this section provide detailed recommendations for specific databases.

Delphix recommends that a separate Database user (i.e. named *Masking User*) be created across all the databases with the appropriate permissions on the schemas to be masked. If needed create multiple users. The appropriate permissions for the database *Masking User* are listed below.

The benefits of having a separate DB *Masking User*:

- Replicating the new user (and privileges) are easier
- Access Audits are much easier
- Can be created as a central AD user and used at many places simultaneously

List of database entitlements required to run masking jobs

- Read data from Tables
- Write data to Tables
- Update data in tables
- Create indexes
- Drop indexes
- Create triggers
- Drop triggers
- Disable triggers
- Enable triggers
- Alter tables add column
- Alter table delete column
- Create constraints
- Delete constraints
- Disable constraints
- Enable constraints

List of database entitlements required to run profiling jobs

- View Definition (Schema)
- Read Data from Tables

Preparing Oracle database for profiling/masking

Overview

Before masking your data, it is important to prepare your database. This article explains the required changes, reasons for the changes, and instructions on how to make the changes.

Archive logging

What is Archive Logging?

Oracle Database lets users save filled groups of redo log files to one or more offline destinations, known collectively as the archived redo log, or more simply the archive log. The process of turning redo log files into archived redo log files is called **archiving**. This process is only possible if the database is running in ARCHIVELOG mode. Users can choose automatic or manual archiving.

Why is it important to make this change?

Archive logging will slow down masking processes and absorb CPU resources that could be used by the masking process. Furthermore, since masking will change every row in every table being masked logs are only needed for short term recovery and transaction backout.

The choice of whether to enable the archiving of filled groups of redo log files depends on the availability and reliability requirements of the application running on the database. If you cannot afford to lose any data in your database in the event of a disk failure, use ARCHIVELOG mode. The archiving of filled redo log files can require you to perform extra administrative operations.

How exactly do I make this change? (exact commands, etc).

```
ALTER DATABASE NOARCHIVELOG;
```

DB/VDB memory allocation

What is SGA? A system global area (SGA) is a group of shared memory structures that contain data and control information for one Oracle database instance. If multiple users are concurrently connected to the same instance, then the data in the instance's SGA is shared among the users. Consequently, the SGA is sometimes called the shared global area.

An SGA and Oracle processes constitute an Oracle instance. Oracle automatically allocates memory for an SGA when you start an instance, and the operating system reclaims the memory when you shut down the instance. Each instance has its own SGA.

The SGA is read/write. All users connected to a multiple-process database instance can read the information contained within the instance's SGA, and several processes write to the SGA during the execution of Oracle. When automatic SGA memory management is enabled, the sizes of the different SGA components are flexible and can adapt to the needs of a workload without requiring any additional configuration. The database automatically distributes the available memory among the various components as required, allowing the system to maximize the use of all available SGA memory. Make sure the DB/VDB memory allocation is sufficient for the workload. Delphix's best practices for sizing a VDB will handle most masking requirements. If you plan to run many concurrent masking jobs a small memory allocation will negatively impact the performance of the masking jobs.

Why is it important to make this change?

To assure that masking jobs will perform at an optimum level.

How exactly do I make this change? (exact commands, etc). Set automatic SGA memory management to enabled. If not allowed set the SGA based on the diagnosis from the AWR report generated during a masking job. The DBA is best suited to make the appropriate tuning changes to the SGA parameters for the version of Oracle being masked.

Undo tablespace size and undo retention time:

What is tablespace? Every Oracle Database must have a method of maintaining information that is used to roll back or undo, changes to the database. Such information consists of records of the actions of transactions, primarily before they are committed. These records are collectively referred to as undo.

Undo records are used to: - Roll back transactions when a ROLLBACK statement is issued - Recover the database - Provide read consistency - Analyze data as of an earlier point in time by using Oracle Flashback Query - Recover from logical corruptions using Oracle Flashback features

When a ROLLBACK statement is issued, undo records are used to undo changes that were made to the database by the uncommitted transaction. During database recovery, undo records are used to undo any uncommitted changes applied from the redo log to the datafiles. Undo records provide read consistency by maintaining the before image of the data for users who are accessing the data at the same time that another user is changing it.

Why is it important to make this change?

The masking Engine updates or inserts masked data in batches. In the case of an insert, it only requires the current transaction size for the commit of each table being masked. The default per table stream is 10k rows. However, with an update, the transaction is not complete until the entire table is masked. So, the more tables and more rows and the wider (size) each row is in each table, the more undo space is needed to complete the transaction. Large tables, such as DW tables or history and Audit tables, most often need an increase to the Undo space and undo Retention time for updates. If space or time is exceeded then the masking job may fail with an ORA-01555, Snapshot too old error.

How exactly do I make this change? (exact commands, etc).

It is highly recommended to increase the Undo space and undo Retention time when running in-place jobs on large tables. A general rule of thumb is 2 or 3 times the size of the largest table(s), or if there are multiple tables running at the same time, then all tables combined. A DBA is best suited to make the necessary UNDO Space and UNDO Retention changes.

Redo logs are optimally sized

What is Redo Logs?

The most crucial structure for recovery operations is the redo log, which consists of two or more preallocated files that store all changes made to the database as they occur. Every instance of an Oracle Database has an associated redo log to protect the database in case of an instance failure.

Why is it important to make this change?

The most important reason to make this change is to keep performance optimal. If redo logs are too small, then the log switching will occur too often, using up valuable Oracle resources.

How exactly do I make this change? (exact commands, etc).

A DBA is best suited to make these changes appropriately.

Change PCTFREE to 40-50:

What is PCTFREE?

PCTFREE and PCTUSED are used together, but PCTFREE is critical for updates. The larger the PCTFREE value the more updates can be done.

Why is it important to make this change?

PCTFREE aids in performance increases for updating Oracle during masking. The Masking Engine does many updates at the same time in batch mode. The more that can be done without DB overhead the faster the masking jobs run.

How exactly do I make this change? (exact commands, etc).

A DBA is best suited to make these changes.

Change primary Key To ROWID:

What is ROWID?

For each row in the database, the ROWID pseudocolumn returns the address of the row. Oracle Database rowid values contain information necessary to locate a row.

Why is it important to make this change?

This is especially important in masking for performance. IF ROWID is used then Oracle will manage the updates for the rows it tracks using ROWID. This makes updates much faster. On occasion, there may be a key (PK/FK/UK) or ID column with an index that is faster, but generally, ROWID is the fastest.

How exactly do I make this change? (exact commands, etc).

Add ROWID as the logical key on each table in the ruleset using the Masking Engine GUI. Also, in a script you should drop foreign keys, and if possible indices and disable triggers and recreate them after the masking job has been run for any of these types of columns being masked.

Masking user privileges:

The following privileges are required for the database user provided in the source connector configuration

- **SELECT** on **V\$CONTROLFILE**
- **SELECT** on **V\$DATAFILE_HEADER**
- **SELECT** on **V\$INSTANCE**

Why is it important to make this change?

These privileges are needed to calculate the size of the source database.

How exactly do I make this change? (exact commands, etc).

SQL commands to grant required privileges to the user

```
GRANT SELECT on v_$controlfile to [mask_user];
GRANT SELECT on v_$datafile_header to [mask_user];
GRANT SELECT on v_$instance to [mask_user];
```

Preparing SQL server database for profiling and masking

Before masking your data, it is important to prepare your database. This section explains the required changes, reasons for the change, and the instructions to make the change.

Logging

What is Simple Recovery Model?

SQL Database Simple Recovery model - Automatically reclaims log space to keep space requirements small, essentially eliminating the need to manage the transaction log space. Operations that require transaction log backups are not supported by the simple recovery model.

Why is it important to make this change?

Reducing the overhead of the transaction logging and the size of the files before checkpoints increases the masking speed significantly.

How exactly do I make this change?

Either (a) use SQL Server Management Studio to open the DB properties dialog box and select the “simple recovery model” or (b) issue the `SET RECOVERY SIMPLE` statement from a SQL query tool. Please see [this reference](#) for more details.

DB/VDB memory allocation

What is min/max memory in SQL Server?

Memory is allocated at the SQL Server level, so all the DBs will share the entire load. The max memory should be close to the maximum available on the server.

Why is it important to make this change?

To assure that masking jobs will perform at an optimum level.

How exactly do I make this change?

Use SQL Server Management Studio and change the max memory allocation for the server.

Primary/Foreign/DMS_ROW_ID Keys

What is a key?

A key is a unique, non-null value that identifies a row in the database.

Why is it important to make this change?

Using a PK or Foreign key is critical for fast updates. When a table does not have an identity column with an index or a PK/FK then the masking engine will alter the table to have an Identity column, DMS_ROW_ID to optimize performance.

How exactly do I make this change?

A logical key can be added to a table in the Masking Engine Ruleset for each table, if there is a specific column that would find the row to update faster than the current PK/FK.

Creating a masking user and privileges

It is highly recommended to create a database user, and possibly a role, for use by the Masking Engine. This user should be created in a non-Production environment and not in your production environment. The following permissions are needed:

- db_datareader
- db_datawriter
- db_ddladmin

SQL commands to add a user with the required privileges:

```
USE [mask_db]
GO
CREATE LOGIN [mask_user] WITH PASSWORD=N'delphix123'
GO
CREATE USER [mask_user] FOR LOGIN [mask_user]
GO
USE [mask_db]
GO
ALTER ROLE [db_datareader] ADD MEMBER [mask_user]
GO
USE [mask_db]
GO
ALTER ROLE [db_datawriter] ADD MEMBER [mask_user]
GO
USE [mask_db]
GO
ALTER ROLE [db_ddladmin] ADD MEMBER [mask_user]
GO
```

Preparing Sybase database for profiling and maskin

- Masking large tables can result in large transactions (depending on the masking job's commit size). It is important to manage each database's transaction log as appropriate to allow the masking jobs to run. Failure to manage the transaction log can result in the suspension of the transaction and hence the masking job appears to hang. Please review the ASE documentation [Managing Free Space with Thresholds](#) on how to manage the transaction log threshold. Sometimes it is necessary to resize the database to have a larger transaction log. When resizing a Delphix VDB, take care to ensure that the any new log devices are created in the VDB's underlying "datafile" directory provided by the Delphix Engine. For more information please review **Resizing an SAP ASE VDB** located on <https://docs.delphix.com/docs>.

Before masking your data, it is important to prepare the database. This section explains the required changes, reasons for the change, and instructions to make the change.

What is min/max memory in SQL server?

Determining the amount of memory SAP ASE needs

The total memory SAP ASE requires to start is the sum of all memory configuration parameters plus the size of the procedure cache plus the size of the buffer cache, where the size of the procedure cache and the size of the buffer cache are expressed in round numbers rather than in percentages. The procedure cache size and buffer cache size do not depend on the total memory you configure. You can configure the procedure cache size and buffer cache size independently. Use **sp_cacheconfig** to obtain information such as the total size of each cache, the number of pools for each cache, the size of each pool, and so on.

Use **sp_configure** to determine the total amount of memory SAP ASE is using at a given moment: `1 >`

```
sp_configure "total logical memory"
```

Parameter Name	Default	Memory Used	Config Value	Run Value	Unit	Type
total logical memory	33792	127550	63775	63775	memory pages(2k)	read-only

The value for the Memory Used column is represented in kilobytes, while the value for the Config Value column is represented in 2K pages.

The Config Value column indicates the total logical memory SAP ASE uses while it is running. The Run Value column shows the total logical memory being consumed by the current SAP ASE configuration. Your output differs when you run this command because no two SAP ASEs are configured exactly the same.

Determine the SAP ASE memory configuration

The total memory allocated during system start-up is the sum of memory required for all the configuration needs of SAP ASE. You can obtain this value from the read-only configuration parameter **total logical memory**. This value is calculated by SAP ASE. The configuration parameter **max memory** must be greater than or equal to **total logical memory**. **Max memory** indicates the amount of memory you will allow for SAP ASE needs.

During server start-up, by default, SAP ASE allocates memory based on the value of **total logical memory**. However, if the configuration parameter **allocate max shared memory** has been set, then the memory allocated will be based on the value of **max memory**. The configuration parameter **allocate max shared memory** enables a system administrator to allocate the maximum memory that is allowed to be used by SAP ASE, during server start-up.

The key points for memory configuration are:

- The system administrator should determine the size of shared memory available to SAP ASE and set **max memory** to this value.
- The configuration parameter **allocate max shared memory** can be turned on during start-up and runtime to allocate all the shared memory up to **max memory** with the least number of shared memory segments. A large number of shared memory segments have the disadvantage of some performance degradation on certain platforms. Check your operating system documentation to determine the optimal number of shared memory segments. Once a shared memory segment is allocated, it cannot be released until the server is restarted.
- The difference between **max memory** and **total logical memory** determines the amount of memory available for the procedure and statement caches, data caches, or other configuration parameters.
- The amount of memory SAP ASE allocates during start-up is determined by either **total logical memory** or **max memory**. If you set **alloc max shared memory** to 1, SAP ASE uses the value for **max memory**.
- If either **total logical memory** or **max memory** is too high:
 - SAP ASE may not start if the physical resources on your machine are not sufficient.
 - If it does start, the operating system page fault rates may rise significantly and the operating system may need to be reconfigured to compensate.

Why is it important to make this change?

To assure that masking jobs will perform at an optimum level.

Primary/Foreign/DMS_ROW_ID keys to for masking Sybase:

What is a key?

A key is a unique, non-null value that identifies a row in the database.

Why is it important to make this change?

Using a PK or Foreign key is critical for fast updates. When a table does not have an identity column with an index or a PK/FK then the masking engine will alter the table to have an Identity column, DMS_ROW_ID to optimize performance.

How exactly do I make this change? (exact commands, etc).

A logical key can be added to a table in the Masking Engine Ruleset for each table, if there is a specific column that would find the row to update faster than the current PK/FK.

Note Sybase ASE will create unavoidable log entries when a table is altered and will increase the log size significantly. If needed, run the masking jobs using the On-The-Fly method to avoid log file increases.



While performing a data copy, the database that contains the table must have select **into/bulkcopy/pllsort** turned on.

Creating a Masking user and privileges:

It is highly recommended to create a database user, and possibly a role, to mask. This user should not be created in production but should be created in non-Production. The following permissions are needed:

Syntax to add user and give privileges:

```
sp_adduser mask_user;  
  
CREATE user NEWUSER;  
  
CREATE LOGIN mask_user WITH PASSWORD Delphix_123; --THIS MUST BE DONE IN MASTER  
  
CREATE USER mask_user IDENTIFIED BY Delphix_123;  
  
GRANT SELECT ON PII_V2 TO mask_user; GRANT INSERT ON PII_V2 TO mask_user; GRANT  
DELETE ON PII_V2 TO mask_user; GRANT ALTER ON PII_V2 TO mask_user; GRANT UPDATE ON  
PII_V2 TO mask_user;  
  
GRANT ALTER ANY TABLE TO mask_user;
```

Adaptive Server requires a two-step process to add a user: sp_addlogin followed by sp_adduser.

```
CREATE LOGIN MASK_SUPER_USER WITH PASSWORD Delphix_123;  
  
sp_addlogin MASK_SUPER_USER, Delphix_123;  
  
GRANT ROLE sa_role TO MASK_SUPER_USER;
```

Connecting data

This section contains the following topics:

- [Managing environments](#)
- [Managing remote mounts for VM continuous compliance engines](#)
- [Managing remote mounts for containerized masking](#)
- [Managing SSL/TLS over JDBC for containerized masking](#)
- [Managing connectors](#)
- [Managing extended connectors](#)
- [Managing rule sets](#)
- [Managing file formats](#)
- [Managing inventories](#)
- [Managing record types](#)
- [Masking whole file](#)
- [JSON file masking](#)

Managing environments

This section describes how you can create and manage your environments in the masking service.

As a reminder, environments are used to group certain sets of objects within the Masking Engine. They can be thought of as folders/containers where a specified user can create manage connectors, rule sets, and jobs.

The Main Environment screen lists all the environments the logged in user has access to. It is the first screen that appears when a user logs into Delphix.

The screenshot displays the 'Environments' management interface. At the top, there's a blue header with 'DELPHIX MASKING' and navigation tabs: 'Environments' (active), 'Monitor', 'Settings', 'Admin', and 'Audit'. A 'Job Wizard' button and a user dropdown 'admin' are also present. Below the header, a breadcrumb 'Home > Environments' and a 'Select Action' dropdown are visible. A search bar with a 'Search' button is provided. The main content is a table with the following data:

Environment ID	Application ▲	Environment	Purpose	No of Jobs	Edit	Export	Copy	Delete
1	test1	test1	Mask	0				

At the bottom of the table, there is a 'Go to top of page' link and a footer with navigation links 'Environments | Monitor | Settings | Admin | Audit' and the 'DELPHIX' logo.

The main **environments** screen contains the following information and actions:

- **Environment ID** — The numeric ID of the environment used to refer to the environment from the Masking API.
- **Application** — A way to indicate the name of the application whose data will be managed within this environment.
- **Environment** — The name of the environment.
- **Purpose** — The purpose of the environment.
- **Jobs** — The number of jobs contained within the environment.
- **Edit** — Edit the environment. See more details below.
- **Export** — Export the environment. See more details below.
- **Copy** — Copy the environment. See more details below.
- **Delete** — Delete the environment. See more details below.

The environments on the screen can be sorted by the various informational fields by clicking on the respective field. In addition, the environments listed can be filtered using the **Search** field. See more details below.

Adding an application

For an environment to be created, an application needs to be specified. Here are the steps to add an application:

The screenshot shows the 'Environments' page in the DELPHIX MASKING application. The page has a blue header with navigation links: Environments, Monitor, Settings, Admin, and Audit. A search bar is located below the header. The main content area displays a table with the following data:

Environment ID	Application	Environment	Purpose	No of Jobs	Edit	Export
1	test1	test1	Mask	0		

A 'Select Action' dropdown menu is open, showing the following options: Add Application (highlighted), Add Environment, Export Settings, Import Settings, Import Environment, and Async Task Status. The page also includes a 'Go to top of page' link and a footer with the DELPHIX logo.

1. On the main environments page, near the upper right-hand corner of the screen, click on the **Select Action** drop-down list and select the **Add Application** option.
2. The screen prompts you for the following items:
 - a. Application Name
3. Click **Save** to return to the **Environments List/Summary** screen.

Creating an environment

Here are the steps you need to take to create an environment:

1. On the main environments page, in the upper right-hand corner of the screen, click on the **Select Action** drop-down list and select the **Add Environment** option.
2. The screen prompts you for the following items:
3. **Application Name** – The name of the application to associate with the environment, for informational purposes.
4. **Environment Name** – The display name of the new environment.
5. **Purpose** – The type of masking workflow for the environment: Mask or Tokenize/Re-Identify.
6. **Enable Approval Workflow** – Whether or not to require approvals of inventories before masking jobs can be run in the environment.
7. Either click **Save** to return to the **Environments List/Summary** screen, or click **Save & View** to display the **Environment Overview** screen.

Exporting settings

To export the Settings:

1. On the main environments page, in the upper right-hand corner of the screen, click on the **Select Action** drop-down list and select the **Export Settings** option.
2. The screen prompts you to take the input for the optional **Passphrase**. You can input the **Passphrase** by clicking the **Use Passphrase** checkbox.
3. Click **Export**.

Export Settings

Use Passphrase

Enter passphrase

Cancel Export

Export Settings

Use Passphrase

Enter passphrase

Cancel Export

All the information related to Settings (Domain, Algorithm, File Format and so on) is exported to a file.

A status pop-up appears. You can wait to finish the download or you can close the download popup page to download the file for later. When the export operation is complete, automatically it will download the export file or you can click on the **Download file** name to download the export file manually. You can also check the export status from [Async Task Status](#) page.

Importing settings

Once you have exported your settings, you can easily import it into another Masking Engine. To import settings:

Import Settings

Use Passphrase Force Overwrite

Enter passphrase

Settings File

Select...

Cancel Import

1. On the main environments page, in the upper right-hand corner of the screen, click on the **Select Action** drop-down list and select the **Import Settings** option.
2. The screen prompts you for the following items:
3. **Passphrase** – You can input the **Passphrase** by clicking the **Use Passphrase** checkbox. If the settings were exported using a passphrase then you must use the same passphrase for the import settings as well otherwise the import operation will fail.
4. **Force Overwrite** – Specify whether the import should fail if an object already exists with the same ID or the existing object should be overwritten. Click on the force overwrite checkbox if you want to overwrite the existing object.
5. **Settings File** – Click on **Select...** button to browse for the exported settings file that contains the information you want to import. (This file must be a previously exported masking environment.)
6. Click **Import** button to start the import operation.

A status pop-up appears. You can wait to finish the import operation or you can close the pop-up page and check the import status for later. When the import operation is complete, it will show the final status of the import operation on the pop-up page. You can also check the import status from [Async Task Status](#) page.

Async task status

To check the async task status:

1. On the main environments page, in the upper right-hand corner of the screen, click on the **Select Action** drop-down list and select the **Async Task Status** option.
2. A pop-up page will appear with the below filter options:
 - a. **Select Task Type** : Select the type to filter the result.
 - b. **Enter Async Task Id** : Enter the Async Task Id to filter the result.
3. Click on **Find** button to find the async task.

Async Task Status

Select Task Type ▾ Enter Async Task Id Find

ID ▾	Type	Status	
814	EXPORT	SUCCEEDED	Download file
813	EXPORT	FAILED	Download log file
812	EXPORT	SUCCEEDED	Download file
811	IMPORT	SUCCEEDED	Passwords and/or SSH keys of connectors need to be updated.
810	EXPORT	SUCCEEDED	Download file
809	IMPORT	FAILED	Download log file
808	IMPORT	SUCCEEDED	Passwords and/or SSH keys of connectors need to be updated.
807	EXPORT	SUCCEEDED	Download file
806	IMPORT	SUCCEEDED	Passwords and/or SSH keys of connectors need to be updated.
805	IMPORT	SUCCEEDED	Passwords and/or SSH keys of connectors need to be updated.

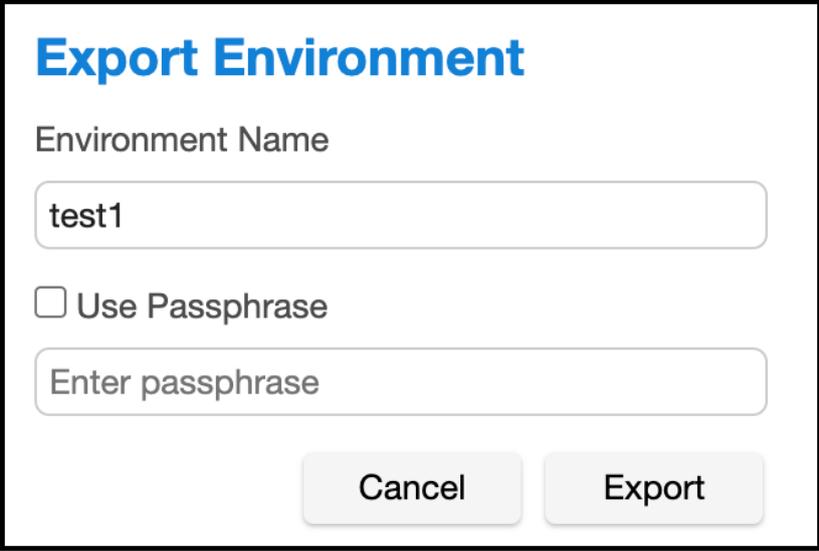
[Cancel](#)

From the result grid, you can also download the export file for the export operation by clicking the **Download file** link on the corresponding row. You can also download the log file for the failed import/export operations by clicking the **Download log file** link on the corresponding row

Exporting an environment

For a variety of different reasons (the main one being moving environments between masking engines), you may want to export all the objects within an environment (connectors, rule sets, masking jobs, etc).

To export an environment use the Export Environment option available in the Masking UI. To export an individual environment:



Export Environment

Environment Name

test1

Use Passphrase

Enter passphrase

Cancel Export

1. Click the **Export** icon or click on **Export** button on the **Environment Overview** screen.
2. The pop-up fills in the following items:
 - a. Environment Name
3. You can input the optional **Passphrase** by clicking the **Use Passphrase** checkbox.
4. Click **Export**.

All the information for the specified environment (connectors, rule sets, inventory, jobs, and so on) is exported to a file.

A status pop-up appears. You can wait to finish the download or you can close the download pop-up page to download the file for later. When the export operation is complete, automatically it will download the export file or you can click on the **Download file** name to download the export file manually. You can also check the export status from [Async Task Status](#) page.

Importing an environment

Once you have exported your environment, you can easily import it into another Masking Engine. To import an environment:

Import Environment

Import Settings Force Overwrite

Application

Existing New

Enter New Application Name

Environment

Existing New

Enter New Environment Name

OTF Environment

Use Passphrase

Enter passphrase

Environment File

Select...

Cancel Import

1. On the main environments page, in the upper right-hand corner of the screen, click on the **Select Action** drop-down list and select the **Import Environment** option.
2. The screen prompts you for the following items:
3. **Import Settings** – Click the checkbox if you want to import settings as well.
4. **Force Overwrite** – Specify whether the import should fail if an object already exists with the same ID or the existing object should be overwritten. Click on force overwrite checkbox if you want to overwrite the existing object.
5. **Application** – You can select the existing application from the application drop-down or you can enter the application name to create a new application.
6. **Environment** – You can select the existing environment from the environment drop-down or you can enter the environment name to create a new environment.

7. **OTF Environment** – Click on **OTF Environment** checkbox to import the on-the-fly connectors into that environment. You can select the existing environment from the environment drop-down or you can enter the environment name to create a new environment.
8. **Passphrase** – You can input the **Passphrase** by clicking the **Use Passphrase** checkbox. If the exported file is used the passphrase then you should use the same passphrase for the import as well.
9. **Settings File** – Click on **Select...** button to browse for the exported settings file that contains the information you want to import. (This file must be a previously exported masking environment.)
10. **Environment File** – Click on **Select...** button to browse for the exported environment file that contains the information you want to import. (This file must be a previously exported Masking environment.)
11. Click **Import** button to start the import operation.

A status pop-up appears. You can wait to finish the import operation or you can close the popup page and check the import status for later. When the import operation is complete, it will show the final status of the import operation on the pop-up page. You can also check the import status from [Async Task Status](#) page.

Editing an environment

To change the properties of an environment, do the following:

1. Click the **Edit** icon to the right of the environment status.
2. The pop-up prompts you for the following information:
 - a. Environment Name
 - b. Purpose
 - c. Application Name
 - d. Enable Approval Workflow
3. Click **Save**.

Copying an environment

A user can also easily create an exact copy of a certain environment. This is a very powerful feature when wanting to have several similar but not exact environments but don't want to start from scratch. To copy an environment do the following:

1. Click the **Copy** icon to the right of the environment status.
2. The pop-up prompts you for the following information:
 - a. Environment Name
 - b. Purpose
 - c. Application Name
 - d. Enable Approval Workflow
3. Click **Save**.

Deleting an environment

To delete an environment:

- Click the **Delete** icon to the right of the environment status and copy icon.



Clicking the **Delete** icon deletes EVERYTHING for that environment: connections, inventory, rule sets, and so on. It does not delete universal settings like algorithms, domains, etc.

Searching for environments

When a large number of environments have been created on a Masking Engine, it may be useful to filter the **Environments List/Summary** screen. To filter the environment list, do the following:

1. In the **Search** field in the upper left side of the screen, enter the characters to search by.
2. Click the adjacent **Search** button.
3. The screen will display only the environments whose name match the specified search characters.

To re-display, the entire list of environments, clear the **Search** field of characters and click the **Search** button again.

Managing remote mounts for VM continuous compliance engines

This section describes how you can mount an NFS/CIFS location inside the Continuous Compliance engine and use it in a masking job for engines deployed on virtual machines. For information on file mounts for containerized masking, please refer to [Managing Remote Mounts for Containerized Masking](#).

In order to access the files shared over NFS/CIFS server from the Masking Engine, complete the following two steps:

1. Create and connect a mount using [Mount Filesystem API](#) endpoint.
2. [Create a file connector](#) with Filesystem Mount Point mode. Or, [Upload a XML/Copybook file format](#) using Filesystem Mount Point mode.

Mount filesystem API

The **Mount Filesystem** APIs are used to perform normal CRUD operations(Create, Read, Update, and Delete) along with three mount operations connect(mount), disconnect(unmount), and remount on a mounted object.

Mount information

To create a mount entry, information about the mount is passed. Some of them are required and some are optional.

- Required Information:
 - **mountName**: The name of the mount. This name is used to refer to this mount in the connector creation and file format upload UIs.
 - **hostAddress**: The NFS/CIFS server address.
 - **mountPath**: The remote path shared by the NFS/CIFS server. For a CIFS mount, this should be the path after the hostname/IP address, with any backslashes (\) replaced with a slash (/). For example, `\10.0.0.1\Share` would be entered as `/Share`.
 - **type**: The type of the server. CIFS, NFS3, or NFS4.
- Optional Information:
 - **options**: The mount options.
 - **connectOnStartup**: Whether this mount should be connected or not when the server starts.

 When a server shuts down, all the mounts are disconnected.

Mount options

The API supports passing many mount options. Not all of them are supported by a server. After a mount is connected, you might see the options field has many options that were not passed by you or some options that have been eliminated that were passed by you. The options field shows effective options only. The applied options are gathered after a mount is connected.

The API also restricts the usage of some mount options.

Enforced options

The following mount options are enforced and added to the list of options for all mounts:

- **nosuid**: The filesystem cannot contain set userid files.
- **noexec**: No executable script can be run from the mount.
- **nodev**: The filesystem cannot contain special devices.

Minimal options

Although `options` is an optional field, it is required for CIFS mounts to pass credentials. The following options are required for CIFS mounts:

- **username:** The username to connect to the CIFS server.
- **password:** The password of the user.
- **domain:** The domain of the user.

For example, `"options": "username=abc,password=pass,domain=DOMAIN"`

For NFSv3 mounts, `options` are not required, therefore can be `null`.

For NFSv4 mounts, the following option is required:

- **nfsvers:** The NFS protocol version number. For example, `"options": "nfsvers=4.0"`

Version options

The version information is passed using `vers` option. The supported versions based on mount types are

Mount Type	Supported Versions
CIFS	2.0, 2.1, 3.0
NFS3	3, 3.0
NFS4	4, 4.0, 4.1, 4.2

Generic options

Some mount options are generic which can be applied to all the mount types while some are mount specific options. In the case of *remount* operation, only generic options can be modified. The list of allowed generic options are:

`async`, `atime`, `auto`, `context`, `defaults`, `defcontext`, `diratime`, `dirsync`, `fscontext`, `group`, `iversion`, `lazytime`, `loud`, `mand`, `_netdev`, `noatime`, `noauto`, `nodev`, `nodiratime`, `noexec`, `nofail`, `noiversion`, `nolazytime`, `nomand`, `norelatime`, `nostrictatime`, `nosuid`, `nouser`, `owner`, `relatime`, `_rnetdev`, `ro`, `rootcontext`, `rw`, `silent`, `strictatime`, `sync`, and `user`.

CRUD operations

Create

The **create** endpoint is used to create a mount entry. It takes all the information about a mount as its input and creates a mount entry. It doesn't do any kind of validation about the mount's accessibility. The validation is done during the *connect* operation.

Read

The **read** endpoints are used to retrieve information about a mount. There are two *read* endpoints.

1. *get all*: To get information about all mounts.

2. *get*: To get information about any particular mount identified by its id.

Update

The **update** endpoint is used to modify any information of a mount. Update operation can be performed only on a disconnected mount.

Delete

The **delete** endpoint is used to delete a mount entry. A mount can be deleted only if it is not being used in any of the connectors.

Mount operations

Apart from normal CRUD operations, there are three special mount related operations exposed through the API.

Connect

The **connect** endpoint is used to mount a remote mount inside the masking engine. If the connect operation succeeds then, the options field is updated with the applied mount options.

Disconnect

The **disconnect** endpoint is used to unmount a remote mount from the Masking Engine.

Remount

The API supports the **remount** operation. This can be used to remount an active or to connect a disconnected mount and also to update some mount information. This can update *mountName*, *connectOnStartup* and generic *options* only. For other updates, use the normal update API.

Resolve mount consistency

A script runs in the background to keep the data in the *mount_information* table and mounts in sync. If for some reason, the data for a mount mounted inside the mount engine and data corresponding to that mount in *mount_information* table becomes inconsistent, the mount is unmounted. For example, if a mount is in a disconnected state in DB but it is mounted in the engine, then it will be unmounted.

Using mounts

A mount can be used at two places:

- File connectors
- File formats

File connector

While creating a connector, when any file connector option is selected, the UI shows a dropdown to select how a file will be accessed. There are three options:

- Filesystem Mount Point
- SFTP
- FTP

Create Connection

Type
File - Delimited

Connection Name

Connection Mode
Connection Mode
Filesystem Mount Point
SFTP
FTP

Cancel Save

On selecting the Filesystem Mount Point option, the mount name and a path inside the mount needs to be specified.

Create Connection

Type
File - Delimited

Connection Name

Mount Name
Choose Mount Name

Connection Mode
Filesystem Mount Point

Path Under Mount
/

Remote Path

Test Connection Cancel Save

- Mount Name: This is a list of mount names created in the engine.
- Path Under Mount: A path relative to the path mounted. By default, it is at the root of the remote Mount path.
- Remote Path: The complete remote path. On selecting a mount name and typing a path in the above input box, this gets updated.

Create Connection

Type
File - Delimited ▼

Connection Name

Mount Name
Choose Mount Name ▼

Connection Mode
Filesystem Mount Point ▼

Path Under Mount

Remote Path

 A connector can be created even if a mount is in a disconnected state but it should be in an active state when a ruleset is being created or when a job is run.

File format

The XML and Copybook file formats can be uploaded from a remote location. To upload a file format from an NFS/CIFS location, select the Filesystem Mount Point option.

Import File Format

Import Format Type
Choose Import Format Type ▼

Import Fields

Sync mounts

A mount can be synced from a source engine to a target engine using [Sync APIs](#). Syncing a file connector using a mount also syncs the related mounts. The following mount information fields are synced:

- mountName
- hostAddress
- mountPath
- options
- connectOnStartup
- type

In case of CIFS mounts, the password is not synced. In order to set the password in the target engine, update the mount's options and ensure to include the password in the options.

Recommended mount server configuration

The NFS and CIFS servers should be configured in such a way that the files are readable and writeable by the Masking Engine.

CIFS server

The user-provided to connect to the mount should have read and write permission on the mount.

NFS server

1. The Masking Engine's server IP should have read and write permission on the mount.
2. For NFS, the access to a file is controlled based on the UID and GID. In order to give read & write permission to the Masking Engine on the share path, the path should be shared with the following options:

```
<mount path> <masking engine ip>(rw,all_squash,anonuid=<uid>,anongid=<gid>)  
# uid and gid is of the owner of the shared path on the server
```

Managing remote mounts for containerized masking

This section describes how to mount an NFS mountpoint inside the Containerized Masking Engine. For information on file mountpoints for Virtual Machine Masking, please refer to [Managing File Mounts](#).

In Containerized Masking, much more control is available to the admin at the Kubernetes layer. That advantage is used to simplify file systems mounts for Containerized Masking. This document will describe the process using NFS as an example mountpoint type.

i Restriction

Filesystem mount points must be mounted as a subdirectory of `/var/delphix/masking/remote-mounts/`.

i Restriction

In order for Kubernetes to utilize some particular network filesystem, the underlying host will typically need to be able to support that filesystem. In this example, to support mounting NFS filesystems, the underlying OS needs to be able to perform an nfs mount. This is typically enabled by installing the nfs-client package. For example, if the kubernetes cluster runs on top of a debian-type linux distro, the package would need to be installed using `apt install nfs-client` on each node to ensure all nodes have the necessary utilities to handle mounting NFS filesystems.

Creating the mountpoint connection in Kubernetes

To establish a remote mount using NFS, the first step is creating the NFS connection to the remote NFS host. This is accomplished utilizing a special NFS persistent volume. This can be added to the beginning of the `kubernetes-config.yaml` file or created as separate config files just for this purpose. If separate config files are created, they will have to be applied before the main Pod config is applied.

Both a Persistent Volume (PV) and Persistent Volume Claim (PVC) are necessary and the YAML for each of these looks like the following snippets.

NFS Persistent Volume YAML

```
apiVersion: v1
kind: PersistentVolume
metadata:
  name: nfs-pv
spec:
  capacity:
    storage: 500Mi
  volumeMode: Filesystem
  accessModes:
    - ReadWriteOnce
  storageClassName: nfs-storage
  mountOptions:
    - hard
    - nfsvers=4.1
```

```
nfs:
  server: <your NFS server host>
  path: <the exported directory on the NFS server, for example /var/tmp/
masking-mount>
```

NFS persistent volume claim YAML

```
apiVersion: v1
kind: PersistentVolumeClaim
metadata:
  name: nfs-pvc
spec:
  accessModes:
    - ReadWriteOnce
  volumeMode: Filesystem
  storageClassName: nfs-storage
  resources:
    requests:
      storage: 500Mi # change corresponding to actual requirements
```

Using the mountpoint in the pod configuration

Next, the recently created NFS PVC must get mounted into the application container. This is achieved by editing the existing Pod config YAML and adding 2 objects. First, attaching the PVC to the pod as a volume. Second, linking that volume into the application container. This is demonstrated in the excerpt below.

Excerpt of kubernetes-config.yaml to show support for NFS volumes

```
#
# Example of volume definition per Persistent Volume Claim
#
volumes:
  - name: nfs-pv-storage
    persistentVolumeClaim:
      claimName: nfs-pvc
containers:
  - image: delphix-masking-app:6.0.16.0-c1
    name: app
    ports:
      - containerPort: 8284
        name: http
    volumeMounts:
      - name: masking-persistent-storage
        mountPath: /var/delphix/masking
        subPath: masking
      - name: masking-persistent-storage
        mountPath: /var/delphix/postgresql
        subPath: postgresql
```

```

#
# Example of mounting an external volume
#
# Mount path is the directory on the `app` container to be mounted to the
# remote provided Persistent Volume.
# It should always start with the `/var/delphix/masking/remote-mounts`
# and to be followed with customer named sub-directory per mount.
# That sub-directory will automatically be created on the Masking Engine
`app` container.
#
- name: nfs-pv-storage
  mountPath: /var/delphix/masking/remote-mounts/nfs_example

```

Using the mountpoint in the UI

Once a properly configured Pod is started, the configured NFS filesystem can be accessed in the UI using the same process that was previously used for non-containerized instances documented in [Managing Remote Mounts for VM Masking Engines](#). The one sticking point is that these mount points (in the dropdown list) by default are named "mountpoint_1", "mountpoint_2", etc.

It is possible to rename the default mount point names to something more friendly. This is done via the `PUT /mount-filesystem/{mountID}` API endpoint.

 The `/mount-filesystem` API has a large set of functionality that is used to manage filesystem mounts in the Virtual Machine deployment of the Masking Engine. For Containerized Masking, most of that functionality is handled by Kubernetes itself rendering the API tasks useless and therefore disabled. The only functionality available in Containerized is the endpoint that allows you to update an existing mount and only to update its name.

Other types of filesystem mountpoint

The above example has used NFS, but it is possible to mount any filesystem that Kubernetes will support. To mount CIFS or some other supported remote filesystem is possible so long as the same general procedure is followed including:

- creating the various Kubernetes objects (such as the PV and PVC)
- mounting it under the `/var/delphix/masking/remote-mounts/` required path

Known limitations

- You can't configure mount point manually (i.e. using API endpoints). Only mount points provided by Kubernetes will be detected.
- Customized mount points can't be synced from Appliance Masking Engine. If the sync bundle contains any mount point created via API - importing that bundle to containerized Masking Engine will fail.
- Masking Sync is incapable of altering your various Kubernetes config YAML files which is the only way to mount a filesystem in Containerized Masking.
- You can't edit existing mount points at containerized Masking Engine.
- Mount points are named automatically by Masking Engine
- You can delete (via API `mountFilesystem`) only those mount points which are not provided by Kubernetes (for example were synced in), and not associated with any existing connector.

Local file masking troubleshooting

If Masking Engine is not responsive at `<your-masking-engine-URL>:30080/masking` - there might be need to troubleshoot. If you are not sure what's the name of the masking pod you can find all pods in the given Kubernetes's cluster by running **`kubectl get pod`** command. The one with the word **masking** will be the desired pod. If multiple masking pods are run on the same instance - look for

`delphix-masking-*` names, Pod status could be seen by running **`kubectl get pod <your-masking-pod-name>`**. If not all 3 containers are in the running status - let's get the description of the pod: **`kubectl describe pod delphix-masking-0`**. In the output of the above command there is the health information for each container, their status and the latest errors that prevented the pod from a successful startup. Most probably those errors will give a hint on what went wrong. If Masking Engine was working fine prior to adding the Local File Masking configuration, the error reasons could be (but not limited to):

- the configured remote Persistent Volume is not accessible
- the directory configured for remote Persistent Volume doesn't exist
- the yaml files entries you've added are not correctly indented (yaml files are indentation sensitive). After fixing the found problem and tearing down all created Kubernetes instances (in the opposite order) - start applying those again.

If Masking Engine application is up and running, but the configured masking job fails - verify the write permissions are granted to the masking target directory (on the corresponding mounted Persistent Volume).

Managing SSL/TLS over JDBC for containerized masking

On the VM instance, we use the Virtualization Engine's Setup App to manage certificates and trust stores for SSL/TLS needs. Since Containerized Masking Engine runs alone - we need to provide another way of creating the truststore and storing the SSL certificate. There are multiple options of establishing truststore on linux container. Below is an example of using Kubernetes for this purpose.

- uploading the saved certificate to configmap
- mounting that configmap as volume
- creating a truststore and uploading there the configured certificates

Prerequisites

Database is configured with SSL listener. To establish the SSL/TLS connection over JDBC we should know:

- database URL,
- SID,
- SSL listener port,
- SERVICE_NAME (for database service where SSL listener is enabled)
- SSL_SERVER_CERT_DN (SSL server certificate distinguished name) - could be found from the generated certificate, for example by using the openssl utility:

```
openssl x509 -in ssl_cert.crt -text
```

Here **ssl_cert.crt** is a name of the file containing the desired certificate (the one that was copied from the Database).

Create configmap entry based on database provided SSL/TLS certificate

1. save SSL/TLS certificate as .crt file.
2. use Kubernetes command to create a configmap, for example:

```
kubectl configmap ora-18 --from-file=ssl_cert.crt
```

Here **ora-18** is the name of the created configmap entry, **ssl_cert.crt** file contains the SSL/TLS certificate. To verify that configmap entry is added to the pod instance run the following command:

```
kubectl get configmap
```

Mount the configured configmap as volume

Add configmap entry as a volume to the pod instance in it's config .yaml file. If you already have other volumes defined that new entry can go under the existing volumes section. If not create a **volumes:** section as shown below:

```
volumes:
  - name: ora-ssl-cert-volume
    configMap:
      name: ora-18
```

Here `ora-ssl-cert-volume` is a name for the provided volume, `ora-18` is the name of the previously created configmap entry.

Now we are ready to mount that volume to **app** container. Under the **containers:** section of the pod's config .yaml file, find the **app** container and add another entry to its **volumeMounts:** as shown below:

```
- name: ora-ssl-cert-volume
  mountPath: /var/delphix/ssl/ssl_cert.crt
  subPath: ssl_cert.crt
```

Here **ora-ssl-cert-volume** is a pod level provided volume, **ssl_cert.crt** is a name of the certificate file (originally provided by the configured configmap).

If using multiple SSL/TLS certificates - the above steps to be repeated for each certificate.



Attention!

The used mountPath `/var/delphix/ssl/` is a preconfigured location on the app container where certificates should be stored! That's where the truststore will look for customer provided certificates.

Create trust store and upload all mounted SSL/TLS certificates

We suggest using Kubernetes's lifecycle `postStart` hook to create the truststore and load the certificates:

In the pod's config .yaml file in the **containers:** section, find the **app** container and add to a lifecycle section to contain a **postStart:** hook as shown below

```
name: app
  lifecycle:
    postStart:
      exec:
        command: ["/bin/bash", "-c", "for filename in /var/delphix/ssl/*.crt;
do keytool -import -trustcacerts -keystore /var/delphix/ssl/.masking_certs -storepass
changeit -noprompt -alias $(basename \"$filename\" .crt) -file \"$filename\"; done"]
```

Here we use the `keytool` utility to create the truststore `/var/delphix/ssl/.masking_certs` and to load all the mounted certificates found in the `/var/delphix/ssl/` directory.

Configure SSL/TLS over JDBC connector

Now any required SSL/TLS certificates are uploaded to the truststore on Containerized Masking Engine. We can use them to establish the JDBC connection. In the connector settings for the advanced Oracle database connector the URL to be configured as following:

```
jdbc:oracle:thin:@(DESCRIPTION=(ADDRESS=(PROTOCOL=tcps)(HOST=<your oracle DB URL>)
(PORT=<port where SSL listener is configured>))(CONNECT_DATA=(SERVICE_NAME=<service
name>))(SECURITY=(SSL_SERVER_CERT_DN="<distinguished name of the SSL certificate>")))
```

SSL/TLS over JDBC troubleshooting

1. verify the file contains the exact SSL/TLS certificate (copied from the DB). It should look like:

```
-----BEGIN CERTIFICATE-----  
MIIBkDCB+gIBADANBgkqhkiG9w0BAQQFADARMQ8wDQYDVQQDEwZiYmRoY3AwHhcNMjIwOTAxMDA0  
...  
uVWk84o=  
-----END CERTIFICATE-----
```

1. verify the certificate is mounted under the correct **/var/delphix/ssl/** directory.
2. verify the certificate is uploaded to the truststore by logging into the bash on the app container and checking truststore exists and how many certificates are loaded:

```
keytool -list -keystore /var/delphix/ssl/.masking_certs -v
```

1. if **app** container didn't start - most probably the mount was not configured correctly. Check the pod description for errors:

```
kubectl describe pod delphix-masking-0
```

Particularly check for indentation issues in the YAML entries because Kubernetes is very sensitive to indentation.

Managing connectors

This section describes how you can create and manage your connectors.

As a reminder, connectors are the way users define the data sources to which the Masking Engine should connect. Connectors are grouped within environments. In order to navigate to the **connectors** screen, click on an environment and then click the **Connector** tab.

DELPHIX MASKING

Job Wizard admin

Environments Monitor Settings Admin Audit

Overview Connector Rule Set Inventory

Home > Environments > test1 > Connector

test1

Create Connection

Connector ID	Connector	Meta Data Source	Type	Edit	Delete
1	mssql	Database	mssql		
2	mysqlserver	Database	mssql		
3	testpost	Database	POSTGRESQL		
1	testfile	File	delimited		

* indicates an extension to included connectors Missing Password / SSH Key

Environments | Monitor | Settings | Admin | Audit

DELPHIX

The **connectors** screen contains the following information and actions:

- **Connector ID** — The numeric ID of the connector used to refer to the connector from the Masking API.
- **Connector** — The name of the connector.
- **Meta Data Source** — The type of connector. One of Database, File, or Mainframe.
- **Type** — The specific type of connector.
- **Edit** — Edit the connector. See more details below.
- **Delete** — Delete the connector. See more details below.

The connectors on the screen can be sorted by the various informational fields by clicking on the respective field.

Creating a connector

To create a new connector:

1. In the upper right-hand corner of the **Connector** tab, click **Create Connection**. The **Create Connection** window appears, prompting you for connection information for the data source you would like to connect to. The required information will change depending on the **Type** of data source you select. For more details on what info is needed to connect to different types (Oracle, AWS RDS, etc) see sections below.
2. Several of our connector types offer two different modes of connecting, **Basic** and **Advanced Mode**. Advanced Mode gives you the ability to specify the exact JDBC URL and add parameters that may not be available in Basic Mode.

Create Connection

Type

Database - Oracle ▼ Basic Advanced

Connection Name **Port**

Schema Name Use Kerberos Authentication

Host Name/ IP **Principal Name**

SID **Password**

Custom Properties File [?](#)

- The fields that appear on the Connector screen are specific to the selected Connector Type (see Connector Types below).
- Click **Save**.

Editing a connector

To edit a connector:

- In the **Connector** tab, click the **Edit** icon for the connector you want to edit.
- Change any information necessary. To change the password:
 - Select the checkbox next to **Change Password**.
 - In the field that appears, enter the new **password**.

Edit Connector: mssql

Database - mssql Basic Advanced

<p>Connection Name</p> <input type="text" value="mssql"/>	<p>Port</p> <input type="text" value="1433"/>
<p>Schema Name</p> <input type="text" value="dbo"/>	<p><input type="checkbox"/> Use Kerberos Authentication</p>
<p>Database Name</p> <input type="text"/>	<p>Login ID</p> <input type="text"/>
<p>Host Name/ IP</p> <input type="text"/>	<p><input type="checkbox"/> Change Password</p>
<p>Instance Name</p> <input type="text" value="MSSQLSERVER"/>	

Custom Properties File ? Select...

Test Connection
Cancel
Save

3. Click **Save**.

Deleting a Connector

To delete a connector, click the **Delete** icon to the far right of the connector name.

i When you delete a connector, you also delete its rule sets and inventory data.

Connector types

Database connectors

The fields that appear are specific to the DBMS Type you select. If you need assistance determining these values, please contact your database administrator.

You can only create connectors for the databases and/or files listed. If your database or file type is not listed here, you cannot create a connector for it.

- **Connection Type**— (Oracle, MS SQL Server, and Sybase only) Choose a connection type:
 - **Basic** — Basic connection information.
 - **Advanced** — The full JDBC connect string including any database parameters.

- **Connection Name** — The name of the database connector (specific for your Delphix application).
- **Schema Name** — The schema that contains the tables that this connector will access.
- **Database Name**— The name of the database to which you are connecting.
Note: The database name field is case-sensitive. It must match exactly with the name of the current database as known to the instance.
- **Host Name/ IP** — The network hostname or IP address of the database server.
- **Use Kerberos Authentication** — (Oracle only, optional) Whether to use Kerberos to authenticate to the database. This box is clear by default. Before Kerberos may be used, the appliance must be properly configured, refer to the [Kerberos configuration instructions](#). If this box is checked, the application authenticates with the Kerberos KDC before connecting to the database, then uses its Kerberos credentials to authenticate to the database instead of a login/password. When Kerberos is enabled, the "Login ID" field is treated as the Kerberos user principal name. The password, if supplied, is used to authenticate the user principal with the KDC. The password field may be left blank if the keytab set during appliance configuration contains keys for the user principal.
Note: Kerberos functionality has been disabled in containerized masking.
- **Login ID** — The user login this connector will use to connect to the database (not applicable for Kerberos Authentication).
- **Password** — The password associated with the Login ID or Username. (This password is stored encrypted.)
- **Use Password Vault**
 - (PostgreSQL, MSSQL, and Oracle only) Whether to use a password vault to authenticate to the database instead of a login ID and password. This box is clear by default. Before a password vault may be used, it must be properly configured. If this box is checked, the selected Credential Path is used to obtain database credentials from the password vault it references.
- **Credential Path**
 - (PostgreSQL, MSSQL, and Oracle only) The path to credentials in a password vault to use for database authentication in lieu of a login ID and password.
- **Principal Name**
 - (Kerberos Authentication only) The name of the Kerberos user principal to use when authenticating with the KDC. The realm portion of the principal may be omitted if it matches the configured default realm.
- **Service Principal**
 - (Sybase with Use Kerberos Authentication only) The name of the Sybase service instance.
- **Port** — The TCP port of the server.
- **SID** — (Oracle only) Oracle System ID (SID).
- **Instance Name** — (MS SQL Server only) The name of the instance. This is optional. If the instance name is specified, the connector ignores the specified "Port" and attempts to connect to the "SQL Server Browser Service" on port 1434 to retrieve the connection information for the SQL Server instance. If the instance name is provided, be sure to make exceptions in the firewall for port 1434 as well as the particular port that the SQL Server instance listens to.
- **Custom Driver Name** — (Generic only) The name of the JDBC driver class, including Java package name.
- **JDBC URL** — (Generic and Advanced connector mode for Oracle, MS SQL Server, and Sybase only) The custom JDBC URL, typically including hostname/IP and port number.
- **Connection Properties File** - A Java properties file to specify configurations for the JDBC connection. See [Database Connection Properties](#) for more information.

All database types have a **Test Connection** button at the bottom left of the New Connector window. We highly recommend that you test your connection before you save it. Do so before you leave this window. When you click **Test Connection**, Delphix uses the information in the form to attempt a database connection. When finished, a status message appears indicating success or failure.

File connectors

The following values appear when any of the file connector types are selected:

- **Connector Name** — The name of the file connector (specific to your Delphix application and unrelated to the file itself).
- **Connection Mode**— Filesystem Mount Point, SFTP, and FTP

F Due to networking complications in containerized masking, FTP is currently disabled in containerized deployments. Delphix is researching options to re-enable FTP (for containerized masking) at a future date.

Create Connection

Type

File - Delimited

Connection Name

Connection Mode

Connection Mode

Connection Mode

Filesystem Mount Point

SFTP

FTP

Cancel Save

The rest of the values appear based on the selected **Connection Mode** value. For **Filesystem Mount Point** connection mode, refer to the corresponding section in the [Managing Remote Mounts](#) page. For other connection modes, the following values appear:

- **Path** — The path to the directory where the file(s) are located.
- **Server Name** — The name of the server used to connect to the file.
- **Port** — The port used to connect to the server.
- **User Name** — The user name to connect to the server.
- **Password** — (non-Public Key Authentication only) The associated password for the server.
- **Public Key Authentication** — (Optional) (Only appears for SFTP.) Check this box to specify a public key. When you check this box, the **Available Keys** drop-down appears. Choose a key from the drop-down. See Delphix Masking APIs for information on uploading public keys to the Masking Engine.

F If you plan to do on-the-fly masking then you will need to create a separate environment and connector to be the source for the files to be masked. The masked files will get put into the directory pointed to by the connector you created previously (the target). However, the file path specified in the connector of the target rule set must point to an existing file the target directory. It does not have to be a copy of the file, just an entry in the directory with the same name. It will be replaced by the masked file.

Starting version 6.0.9.0 the SFTP mode is extended with the 'User Directory as root' flag. If the Path defined is relative to the User-home-dir as configured on the SFTP Server, tick the flag below.

Create Connection

Type
File - Delimited ▼

Public Key Authentication

Connection Name

Path

User Directory as root

Connection Mode
SFTP ▼

User Name

Server Name

Password

Port

Test Connection Cancel Save

If the connector is configured via the API then that flag is accessible as "userDirIsRoot", for example:

```
{
  "connectorName": "Test SFTP Connector",
  "environmentId": 2,
  "fileType": "DELIMITED",
  "connectionInfo": {
    "connectionMode": "SFTP",
    "path": "/delimited",
    "host": "yourSFTPServer",
    "loginName": "xxxxx",
    "password": "xxxxx",
    "port": 22,
    "userDirIsRoot": true
  }
}
```

Database connection properties

Getting properties

To retrieve all properties set on the connector, make a request to the `GET database-connector/{id}/properties` endpoint. This endpoint will respond with all default properties set by the driver, superimposed by any properties specified by an uploaded connection properties file. If a properties file is uploaded for a connector, this list can also be viewed through the UI on the database connector form, where you can sort by `Property`, `Value`, or `Modified`. The `Modified` field signifies whether the property value is the default or modified by the uploaded properties file.

- The database name field is case-sensitive. It must match exactly with the name of the current database as known to the instance.
- Only a valid JDBC URL is required to retrieve properties of a connector; a valid connection to the database server is not necessarily required.

Edit Connector: mssql

Database - mssql Basic Advanced

Connection Name	Port
<input type="text" value="mssql"/>	<input type="text" value="1433"/>
Schema Name	<input type="checkbox"/> Use Kerberos Authentication
<input type="text" value="dbo"/>	Login ID
Database Name	<input type="text" value="user"/>
<input type="text" value="db"/>	<input type="checkbox"/> Change Password
Host Name/ IP	
<input type="text" value="mysql-server.test.co"/>	
Instance Name	
<input type="text" value="MSSQLSERVER"/>	

Custom Properties File [?](#)

Setting Properties

Properties can sometimes be set through the JDBC URL or through a connection properties file. Customizing the JDBC URL is limited to Advanced, Generic, and Extended Connectors, while uploading a properties file is supported by all database connectors. All properties files must have the extension `.properties` and must adhere to Java properties file syntax. Even if a property specified in the properties file is not technically supported by the JDBC driver, it will still be passed along to the driver when building the JDBC Connection. All provided and unsupported properties will be logged whenever the properties file is loaded.

 The properties file is assumed to be written using ISO 8859-1 character encoding

If possible, specify sensitive properties through relevant form fields which will be obfuscated in all places or through the JDBC URL which will still be visible in plain text to any user with the `VIEW connector` privilege but will be redacted in support bundles.

Managing extended connectors

Extended Connectors allow you to upload additional JDBC Drivers to the Continuous Compliance Engine to enable masking of data sources not natively supported by Continuous Compliance.

Limitations

Delphix supports type 4 JDBC Drivers. These must be a pure-java .jar file that can be used simply by uploading it (or it's zip file) to the engine. Anything that requires compilation on the engine, or execution of any kind of install or licensing script, is not supported.

Extended Connectors don't support all of the features available for built-in connectors like Oracle. As of 6.0.9.0, the "Disable Constraint", "Disable Trigger" and "Drop Indexes" options can be implemented and enabled by driver support plugins, which are detailed [here](#). Delphix provides support for Extended Connectors in accordance with our [Support Policy](#).

Drivers that require a Java version higher than 8 are not supported.

Installing a new driver

To use a new JDBC driver, first you need to upload it to your Masking Engine. Since some drivers require multiple files, the driver and any additional files it needs to function should be put together in a single zip file. Even if a driver doesn't require additional files, it still needs to be zipped.

For example, to package the Informix JDBC driver for use with Continuous Compliance take all three files provided for Informix and zip them together:

```
$ ls
LICENSE.txt ifxjdbc.jar ifxlang.jar
$ zip informix.zip *
  adding: LICENSE.txt (deflated 70%)
  adding: ifxjdbc.jar (deflated 4%)
  adding: ifxlang.jar (deflated 4%)
$ ls
LICENSE.txt ifxjdbc.jar ifxlang.jar informix.zip
$
```

To upload the driver package to the engine, navigate to the **JDBC Drivers** under **Settings**.

Home > Settings > JDBC Drivers

Settings

Add JDBC Driver

JDBC Drivers

Note: Extensions support policy is defined [here](#)

- Algorithms
- Domains
- Profile Sets
- Classifiers
- Expressions
- Roles
- File Formats
- JDBC Drivers**

C							Displaying 1 to 6 of 6
Name	Class Name	Version	Date	Description	Checksum	Actions	
Oracle	oracle.jdbc.OracleDriver	19.3	1 May 2023 12:17 IST	The default Oracle driver used for Oracle connectors.		...	
MSSQL	com.microsoft.sqlserver.jdbc...	8.4.1	1 May 2023 12:17 IST	The default MSSQL driver used for MSSQL connectors.		...	
MySQL	org.mariadb.jdbc.Driver	2.7.2	1 May 2023 12:17 IST	The default MySQL driver used for MySQL and MariaDB connectors.		...	
Sybase	com.sybase.jdbc42.jdbc.Syb...	16.0	1 May 2023 12:17 IST	The default Sybase driver used for Sybase connectors.		...	
Postgres	org.postgresql.Driver	42.5.4	1 May 2023 12:17 IST	The default Postgres driver used for Postgres connectors.		...	
DB2	com.ibm.db2.jcc.DB2Driver	4.25	1 May 2023 12:17 IST	The default DB2 driver used for DB2 connectors.		...	

Environments | Monitor | Settings | Admin | Audit

Support

Clicking **Add Driver** will bring up a dialog box to upload the driver zip file and enter the driver's configuration details.

Add JDBC Driver

Name

Description (Optional)

Class Name

(This is the class that implements java.sql.Driver)

Driver ⓘ
Select the JDBC driver for upload

Extensions Support Policy is defined [here](#)

The **Add Driver** screen lets you set the following information.

- **Name** A human-readable name for the driver. Name it whatever is convenient for you. **Note:** Special Characters are not allowed in the Name field.
- **Description** A human-readable description of the driver.
- **Class Name** The Fully Qualified Class Name of the class in the JDBC driver that implements the java.sql.Driver interface. The class name will be in the documentation for the driver itself.
- **Select JDBC driver for upload** Lets you select the zip file containing the driver and upload it.

 Users cannot update the driver support that a JDBC driver references or uses via the UI; as of now, that can only be done via the web API.

To remove an uploaded driver, click the Actions button to the right side corner of the **JDBC Drivers** list and select the option **Delete**. Note that the delete will fail if any Connectors exist that use the driver you're trying to delete.

If you find you need to edit a driver's configuration options later, click the Actions button to the right side corner of the **JDBC Drivers** list and select the option **Edit**.

Driver permissions

The Continuous Compliance Engine uses the Java Security Manager to prevent uploaded JDBC drivers from performing certain actions without your permission.

Uploaded drivers are granted all permissions *except* for the following non- `FilePermission` :

Class	Target	Action
<code>java.net.SocketPermission</code>	<code>localhost:-</code>	accept, connect, listen, resolve
<code>java.lang.RuntimePermission</code>	<code>exitVM</code>	
<code>java.lang.RuntimePermission</code>	<code>createClassLoader</code>	
<code>java.lang.RuntimePermission</code>	<code>accessClassInPackage.sun</code>	
<code>java.lang.RuntimePermission</code>	<code>setSecurityManager</code>	
<code>java.security.SecurityPermission</code>	<code>setPolicy</code>	
<code>java.security.SecurityPermission</code>	<code>setProperty.package.access</code>	

With regards to `FilePermissions`, `read` access is granted to all, though `write` is only allowed for the following directories:

- the masking user's home directory (`System.getProperty("user.home")`)
- the JVM's default temp directory (`System.getProperty("java.io.tmpdir")`)

Please note that both of these locations are shared, so care will need to be taken to avoid collisions.

The set of permissions granted to uploaded drivers is static and cannot be modified.

Extended logging

The Continuous Compliance Engine provides enhanced logging for extended connectors to assist in debugging connection problems. Enhanced logging can be enabled when the connector is created by checking the 'Enable Logger' box. Enhanced logging may have an impact on performance so you should enable it only when debugging connection problems.

Note that extended logging will not work with signed drivers such as MSSQL.

Enhanced Logging requires some additional permissions to be granted.

Class Name	Target Name	Action Name	Purpose
java.io.RuntimePermissio n	getClassLoader		Allows the driver to load the classes implementing the logging feature

Creating an extended connector

Creating a connector using an Extended Driver is very similar to creating a connector with built-in support. Choose **Database - Extended** as the Type. The following fields will be available:

- **Connection Name** A name for this connection
- **JDBC Driver** Select the JDBC Driver you want to use for this connection
- **Login ID** The username the Masking Engine should connect to the target database with.
- **Password** The password to use to connect to the database
- **JDBC URL** You must provide the JDBC URL for the database to connect to. The exact format and available parameters are specific to the database you're connecting to. Consult your database vendors documentation for details.



Some databases allow you to specify usernames and passwords in the JDBC URL. It's best not to do this. The Continuous Compliance Engine is careful not to log the Login ID and Password in the Masking Engine's logs, but JDBC URLs may be logged unmodified.

Create Connection

Type

Database - Extended Enable Logger

Connection Name **Login ID**

My Informix Connection dbuser

Schema Name **Password**

Informix

JDBC Driver

Choose JDBC Driver

JDBC URL

jdbc:informiz-sqli://informixdatabase.testing.delphix.com:9088/sysuser:INFO|RMIXSES

Type: (none)
Version: (none)
Date Uploaded: (none)
Uploaded By: (none)
Description: (none)

Note: Extensions Support Policy is defined [here](#)

Custom Properties File ?

Once the connector is created, you can create rulesets, inventories, and jobs to profile and mask your data as with other types of connectors.

Extended Connectors can be edited and deleted in the same way as [Built In Connectors](#).

Synchronization

Connectors using extended JDBC Drivers can be synchronized similarly to other connectors. See [Working with Multiple Masking Engines](#) for details. When a job or connector requires an uploaded JDBC Driver, the driver will be

exported along with the connector or job. JDBC Drivers are part of the **Global Object** and so will be synchronized whenever the Global Object is synchronized. They can also be synchronized individually.

License entitlement for commercial JDBC drivers

Continuous Compliance requires the installation of a Delphix license to use certain commercial JDBC drivers. Users who have purchased this entitlement will have access to their given license files on the [Delphix download site](#).

Installing a license



- Commercial JDBC drivers from the following vendors must be licensed using this mechanism:
 - CData
- To synchronize License files attached to a JDBC driver while synchronizing a connector, Global Objects should be Synchronized first.

1. Download the license from the Delphix download site.
2. Upload the license to Continuous Compliance using either
 - a. the API using the `POST /license` API endpoint or
 - b. the UI using the **License** file upload button in **the Add or Edit JDBC Driver** dialogue box. The License button appears on entering a valid CData JDBC driver class name as shown in this screenshot:

Add JDBC Driver

Name

Description (Optional)

Class Name

`cdata.idbc.`

(This is the class that implements java.sql.Driver)

Driver ⓘ

Select the JDBC driver for upload

Choose file

License ⓘ

If you have received a license from Delphix for this driver, select it for upload

Choose file

Extensions Support Policy is defined [here](#)

Cancel

Save

Managing licenses

To view the licenses installed on an engine, use the `GET /license` API or `GET /ALL` API endpoint.

Managing rule sets

This section describes how rule sets can be created, edited, and removed.

The rule sets screen

From anywhere within an Environment, click the **rule set** tab to display the rule sets associated with that environment. The **rule sets** screen appears. If you have not yet created any rule sets, the rule set list is empty.

DELPHIX MASKING

Job Wizard admin

Environments Monitor Settings Admin Audit

Overview Connector Rule Set Inventory

Home > Environments > test1 > Rule Set

Rule Set

Search Search

Create Rule Set

Rule Set ID	Name	Meta Data Source	Type	Edit	Refresh/Save	Copy	Delete
2	fileconnector	File	delimitedFile		N/A		
1	test	Database	mssql				

Go to top of page

Environments | Monitor | Settings | Admin | Audit

DELPHIX

The **rule sets** screen contains the following information and actions:

- **rule set ID** — The numeric ID of the rule set used to refer to the rule set from the Masking API.
- **Name** — The name of the rule set.
- **Meta Data Source** — The type of rule set. One of Database, File, or Mainframe.
- **Type** — The specific type of ruleset.
- **Edit** — Edit the rule set. See more details below.
- **Refresh/Save** — Refresh the rule set. Only applies to Database rule sets. See more details below.
- **Copy** — Copy the rule set. See more details below.
- **Delete** — Delete the rule set. See more details below.

The rule sets on the screen can be sorted by the various informational fields by clicking on the respective field.

The create/Edit rule set window

In the upper right-hand corner, click the **Create rule set** button.

The **Create rule set** window appears.

Create Rule Set

Pick a connector to list its Tables/Files. Check one or more Tables/Files to select them for inclusion in the Rule Set. To remove the Table/file, deselect it.

Name 1
test1

Connector 2
testfile

File Name Patterns 9
Add Pattern 10

Use regular expression to match any files. 11 12

Search Use * to match any characters. 5 Clear 6

Selected: 0 4

- sample.xml
- example.xml
- example_1.xml
- example_2.xml
- example_3.xml

3

7 8 Cancel Save

1	<p>Rule Set Name Input Field</p> <p>When editing an existing rule set, this field will be filled with the existing rule set name by default.</p>
2	<p>Connector List</p> <p>When creating a new rule set, all available connectors will be listed here. When editing an existing rule set, only the connector currently in use will appear.</p>
3	<p>Table or File List</p> <p>If a database connector is selected in the connector list, all available tables in the database schema associated with the connector will appear in this list. If a file connector is selected, all available files in the directory associated with the connector will appear in this list.</p>
4	<p>Selected Table or File Number</p> <p>Displays how many tables or files you have selected.</p>

5	<p>Search Query Input Field</p> <p>You can enter a search query here. After typing the search query, press ENTER to execute the search query.</p> <div data-bbox="252 412 1423 703" style="background-color: #e6e6ff; padding: 10px;"> <p> search query</p> <ul style="list-style-type: none"> • Use * to match any characters in the names of tables or files. • If you have selected a table or file before searching and it is not in the search results, it will not be included in the rule set. You can add back the table or file by removing the search query. • Checkbox / selections do not persist through a search or a clearing of the search field. </div>
6	<p>Clear Search Button</p> <p>Click to remove any search query.</p>
7	<p>Select All Button</p> <p>Click to select all tables or files in the table or file list.</p>
8	<p>Clear All Button</p> <p>Click to deselect all tables or files in the table or file list.</p>
9	<p>File Name Patterns Editor</p> <p>This editor will appear only when the selected connector is a file connector.</p>
10	<p>Add File Pattern Button</p> <p>Click to add a new file pattern entry below.</p>
11	<p>File Pattern Input Field</p> <p>Enter the file pattern here.</p> <div data-bbox="252 1476 1423 1682" style="background-color: #e6e6ff; padding: 10px;"> <p> file pattern syntax</p> <p>Expressions are case sensitive. A file pattern uses the regular expression syntax defined by the Java Pattern class. The syntax is documented here. For example, the pattern <code>.*\.txt</code> will match any file with a <code>.txt</code> extension, such as <code>example.txt</code>.</p> </div>
12	<p>Remove File Pattern Button</p> <p>Click to remove a file pattern.</p>

Creating a rule set

To create a new rule set:

1. Click on the name of an Environment, and then click the **rule set** tab.
2. In the upper right-hand corner of the **rule set** screen, click **Create rule set**.
3. The **Create rule set** screen lets you specify which tables belong in the rule set.
4. Enter a **name** for the new rule set.
5. Select a **Connector** name from the drop-down menu.
6. The list of tables for that connector appears. If you have not yet created any connectors, the list is empty. Click individual table names to select them, or click **Select All** to select all the tables in the connector. See "Create/Edit rule set Window" for a description of the screen and other options.
7. Click **Save**.

You may then need to define the rule set by modifying the table settings as described in "Modifying Tables in a rule set" below.

For example:

- For a table in a database rule set, you may want to filter data from the table.
- For a file in a file or mainframe rule set, you must select a File Format to use.

Refreshing a rule set

Refreshing a rule set will result in the columns in the tables in the rule set being rescanned. As a result, the inventory associated with the rule set will also be refreshed, but any pre-existing algorithm assignments will be retained.

To refresh a rule set:

1. Click the **Refresh/Save** icon to the right of the rule set on the **rule set** screen.
2. The **Refresh/Save** icon will turn to an hourglass as the associated tables are rescanned.
3. After the refresh is complete, the **Refresh/Save** icon will return to the circular arrow.

Copying a rule set

If you copy a rule set, the inventory associated with that rule set will also be copied. Also, any filter conditions defined for that rule set will be copied.

To copy a rule set:

1. Click the **Copy** icon to the right of the rule set on the **rule set** screen.
2. The **Copy rule set** window appears.
3. Enter a **Name** for the new rule set.
4. Click **Save**.
5. Modify the rule set as you want, using the procedures described above.

Deleting a rule set

If you delete a rule set, the inventory associated with that rule set will also be deleted. Also, any filter conditions defined for that rule set will be deleted.

To delete a rule set, click the **Delete** icon to the right of the rule set on the **rule set** screen.

The rule set screen

From the **rule set** tab, click on a rule set to display the tables or files in the rule set. The **rule set** screen appears.

DELPHIX MASKING

Job Wizard admin

Environments Monitor Settings Admin Audit

Overview Connector **Rule Set** Inventory

Home > Environments > test1 > Rule Set

Rule Set Create Rule Set

Search Search

Rule Set ID	Name	Meta Data Source	Type	Edit	Refresh/Save	Copy	Delete
2	fileconnector	File	delimitedFile		N/A		
1	test	Database	mssql				

[Go to top of page](#)

Environments | Monitor | Settings | Admin | Audit

DELPHIX

The **rule set** screen contains the following information and actions:

- **Table or File or Pattern** — The name of the table or file/file pattern in the rule set.
- **Edit** — Edit the table or file in the rule set. See more details below.
- **Delete** — Delete the table or file from the rule set.

For rule sets with a large number of tables or files, the **rule set** screen will be displayed on pages that can be navigated by the controls at the bottom of the list on the page. The tables or files displayed may also be filtered using the **Search** field and button.

Editing/Modifying a rule set

To edit a rule set:

1. Click the **Edit** icon to the right of the rule set on the rule set screen.
2. Click the **Edit rule set** button towards the top.
3. The **Create rule set** screen appears. This screen lets you specify which tables belong in the rule set.
4. Modify the rule set as you want, using the preceding procedures.

Removing a table or File

To remove a table or file from a rule set:

1. From the **rule set** screen, click the **name** of the desired rule set.
2. Click the red **delete** icon to the right of the table or file you want to remove.

Modifying tables in a rule set

If you remove a table/file from a rule set and that table/file has an inventory, that inventory will also be removed.

The features in this section are disabled for file and mainframe rule sets.

You can modify tables in a rule set as follows:

Logical key

A logical key is a unique, non-null value that identifies a row in the database.

If your table has no primary keys defined in the database, and you are using an In-Place strategy, you must specify an existing column or columns to be a logical key. This logical key does not change the target database; it only provides information to Delphix. For multiple columns, separate each column using a comma. Note: If no primary key is defined and a logical key is not defined an identity column will be created.

To enter a logical key:

1. From the **rule set** screen, click the **name** of the desired rule set.
2. Click the green **edit** icon to the right of the table whose filter you wish to edit.
3. On the left, select **Logical Key**.
4. Edit the text for this property. The logical key cannot be more than 1024 characters in length.
5. To remove any existing code, click **Delete**.
6. Click **Save**.

Edit filter

Use this function to specify a filter to run on the data before loading it to the target database.

To add a filter to a database rule set table or edit a filter:

1. From the **rule set** screen, click the **name** of the desired rule set.
2. Click the green **edit** icon to the right of the table you want.
3. On the left, select **Edit Filter**.
4. Edit the properties of this filter by entering or changing values in the **Where** field.

Be sure to specify column name with table name prefix (for example, customer.cust_id <1000).

1. To remove an existing filter, click **Delete**.
2. Click **Save**.

Custom SQL

Use this function to supply a customized SQL SELECT Query for the table. Typically, this query will include a **WHERE** clause to filter or subset the data.

 The custom SQL must contain the primary key column (or columns if the table uses a composite primary key) and all columns that will be masked.

To add or edit SQL code:

1. From the **rule set** screen, click the **name** of the desired rule set.
2. Click the green **edit** icon to the right of the table you want.
3. On the left, select **Custom SQL**.
4. Enter the custom SQL code for this table.

Delphix will run the query to subset the table based on the SQL you specify.

1. To remove any existing code, click **Delete**.
2. Click **Save**.

Creating a ruleset for file formats

Once you create a ruleset with a file or set of files, you will need to assign those files to their appropriate file format.

This is accomplished by editing the ruleset. Click on the edit button for the file the Edit File window will appear with the file name. From the format drop-down select the proper format for the file.

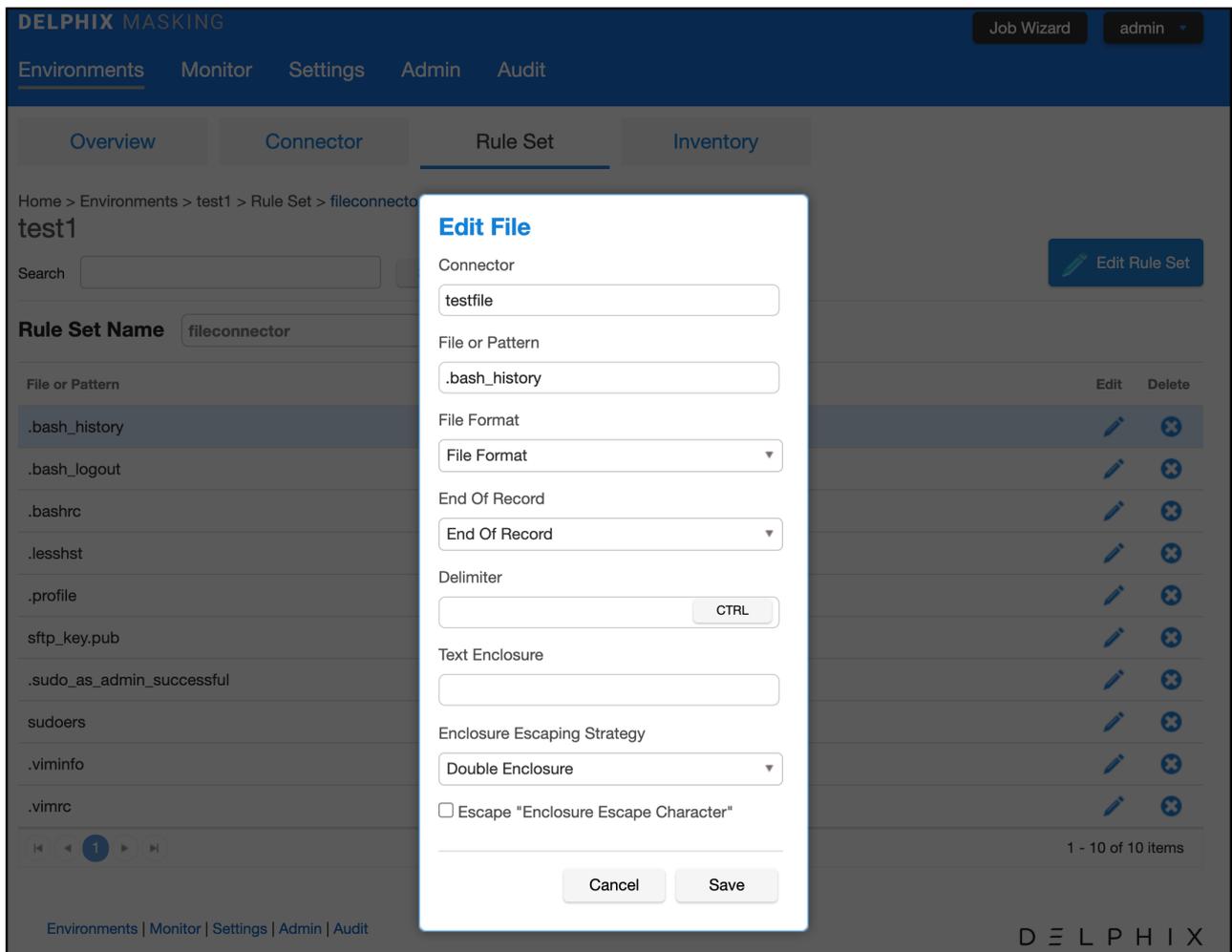
- If the file is a Mainframe data sets file with a copybook you will see a checkbox to signify if the file is variable length.
- For all other file types, select the end-of-record to let Delphix know whether the file is in windows/dos format (CR+LF) or Linux format (LF).
- If the file is a delimited file you will have a space to put in the delimiter.
- If there are multiple files in the ruleset you will have to edit each one individually and assign it to the appropriate file format.

Control character support for delimited files

The user can specify control character as a delimiter/end of record from UI/API.

Control Character

The control character value from UI/API should be in **\$(hex value of the control character)** format , like **\$(01)** for ^A. The control character value support UTF-8 character set.



Control character as a delimiter

1. In order to use control character as a delimiter, the user needs to click on **CTRL** button inside delimiter input text.
2. Clicking on **CTRL** button will open a virtual keyboard where users can select the required control character. Also if the user wants to enter the control character manually then they can use the given format **\$(hex value of the control character)** , like `$(01)` for `^A`.

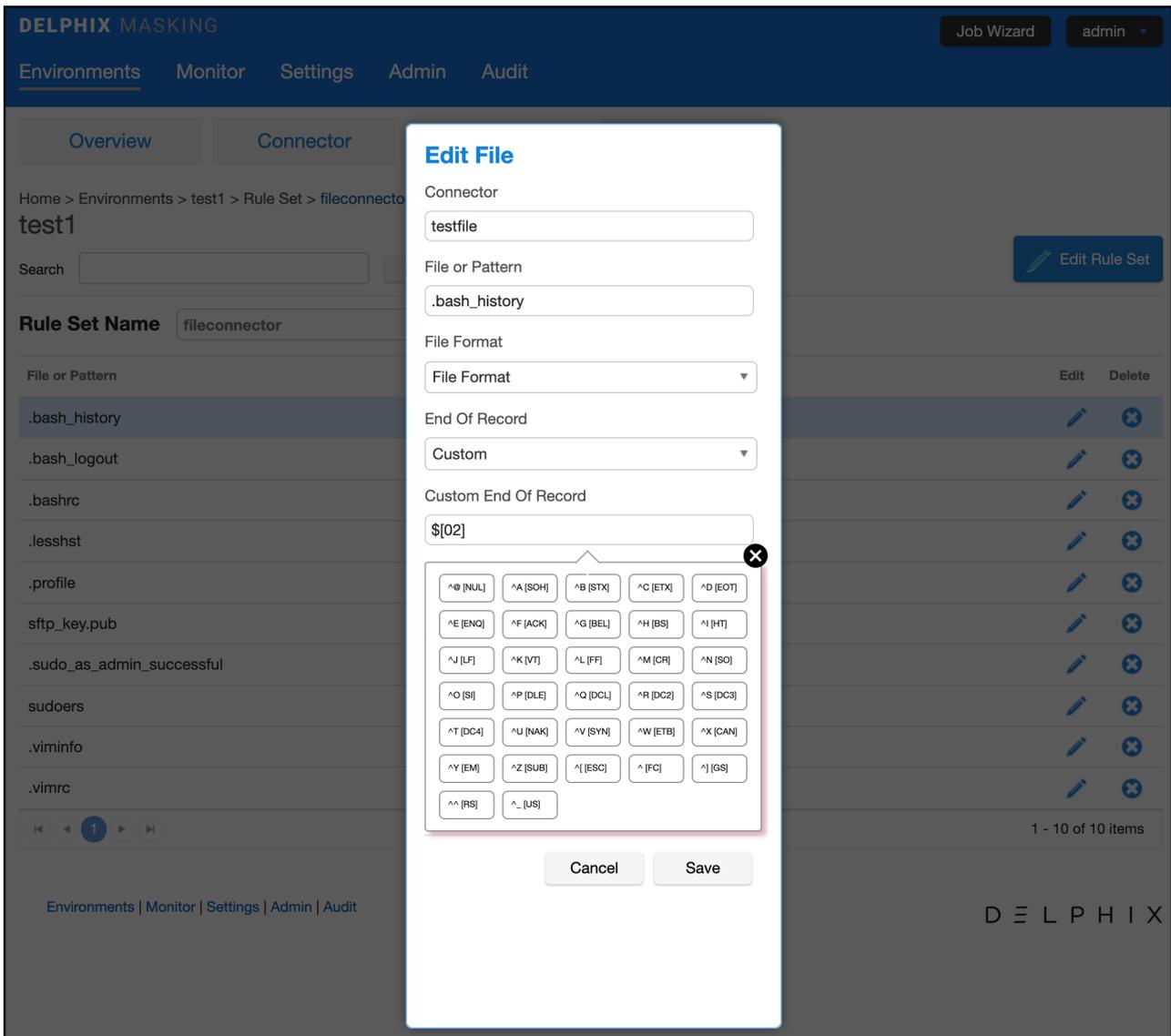
The screenshot displays the 'DELPHIX MASKING' application interface. The main navigation bar includes 'Environments', 'Monitor', 'Settings', 'Admin', and 'Audit'. The user is logged in as 'admin'. The current view is 'Connector' for a rule set named 'fileconnector'. The 'Edit File' modal is open, showing the following fields:

- Connector: testfile
- File or Pattern: .bash_history
- File Format: File Format
- End Of Record: Custom
- Custom End Of Record: `$(02)` (with a CTRL button)
- Delimiter: `$(03)` (with a CTRL button)

A virtual keyboard is open below the Delimiter field, displaying various control characters such as `^@ [NUL]`, `^A [SOH]`, `^B [STX]`, `^C [ETX]`, `^D [EOT]`, `^E [ENQ]`, `^F [ACK]`, `^G [BEL]`, `^H [BS]`, `^I [HT]`, `^J [LF]`, `^K [VT]`, `^L [FF]`, `^M [CR]`, `^N [SO]`, `^O [SI]`, `^P [DLE]`, `^Q [DCL]`, `^R [DC2]`, `^S [DC3]`, `^T [DC4]`, `^U [NAK]`, `^V [SYN]`, `^W [ETB]`, `^X [CAN]`, `^Y [EM]`, `^Z [SUB]`, `^[[ESC]`, `^F [FC]`, `^_ [GS]`, `^^ [RS]`, and `^_ [US]`. The background shows a list of files with 'Edit' and 'Delete' actions for each.

Control character as an end of record

1. In order to use control character as an end of record, the user needs to click on **CTRL** button inside custom end of record input text.
2. Clicking on **CTRL** button will open a virtual keyboard where users can select the required control character. Also if the user wants to enter the control character manually then they can use the given format **\$(hex value of the control character)** , like `$(01)` for `^A`.



Control character as a value

1. Control characters are supported as values in a delimited file. No special configuration is necessary. Simply configure the delimited file format as usual.
2. The user doesn't need to configure anything extra if the control character is only part of the value and not being used as a delimiter or end of record. However, the user needs to define delimiter/end of record as per the requirement.

Define enclosure escaping strategy for delimited files

The user can configure the enclosure escape character from the UI/API to escape the enclosure. To configure the enclosure escape character from the UI, user needs to select the "Enclosure Escaping Strategy" dropdown value as per below options on the edit ruleset popup window,

Double enclosure

Double enclosure option will set the escape character value same as enclosure value. For example, if the enclosure escape character is " then escape character value will be " as well.

The screenshot shows the DELPHIX MASKING interface with the 'Rule Set' tab selected. A modal dialog titled 'Edit File' is open, displaying the following configuration for a file pattern:

- Connector: testfile
- File or Pattern: .bash_history
- File Format: File Format
- End Of Record: End Of Record
- Delimiter: | (CTRL)
- Text Enclosure: *
- Enclosure Escaping Strategy: Double Enclosure
- Escape "Enclosure Escape Character"

The background interface shows a list of rule sets with columns for 'Edit' and 'Delete'. The current rule set is 'fileconnector'.

Custom

By selecting custom option user can specify any single character as an enclosure escape character except the "escape sequences" and "control characters".

The screenshot shows the 'Edit File' dialog in the DELPHIX MASKING interface. The dialog is titled 'Edit File' and is used to configure a specific file within a rule set. The background shows the 'Rule Set' tab with a table of files and their actions.

File or Pattern	Edit	Delete
.bash_history		
.bash_logout		
.bashrc		
.lesshst		
.profile		
sftp_key.pub		
.sudo_as_admin_successful		
sudoers		
.viminfo		
.vimrc		

Default Enclosure Escape Character

The default value for "Enclosure Escaping Strategy" is "Double Enclosure".

Escape "enclosure escape character"

Selecting this checkbox indicates whether the enclosure escape character also escapes itself. For example, if the enclosure escape character is " then the sequence "" would be treated as a single " character, rather than an escape.

Configure enclosure escape character for the large ruleset

To configure the enclosure escape character for the large ruleset user can use this [API Script](#).

Managing file formats

File formats

Unlike database files for the most part do not have built-in metadata to describe the format of the fields in the file. You must provide this to Delphix so it can update the file appropriately. This is done through the settings tab where you will see a menu item on the left for File Format. Select File Format and you will see an option to import a file format. This will depend on the type of file and how you want to let Delphix know the format of the file.

The screenshot shows the Delphix Masking web interface. The top navigation bar includes 'Job Wizard' and 'admin'. The main navigation menu has 'Environments', 'Monitor', 'Settings' (selected), 'Admin', and 'Audit'. The breadcrumb trail is 'Home > Settings > File Format'. The page title is 'Settings' with an 'Import Format' button. The 'File Formats' section features a left-hand menu with options: Algorithms, Custom Algorithms (legacy), Domains, Profiler, Roles, File Formats (selected), and JDBC Drivers. The main content area contains a table with columns: ID, Name, Type, and Delete. The table is currently empty. The footer includes a navigation trail 'Environments | Monitor | Settings | Admin | Audit' and the 'DELPHIX' logo.

Mainframe data sets and XML files

For Mainframe data sets, you can specify the file format via the Import Format button which will import the copybook directly into Delphix. You can input this file from a Filesystem Mount Point, SFTP server, FTP server, or via upload. Please select Copybook as the Import Format Type.

For XML files you can also import the file format with the input format option which will import the file directly into Delphix. You can use the file you want to mask as the format. You can input this file from a Filesystem Mount Point, SFTP server, FTP server, or via upload. Please select XML as the Import Format Type.

Delimited and fixed files

For Delimited and Fixed files you can import a text file that describes the structure of the file to Delphix.

To input the file format for delimited files, create a text document with the column names each on its own line. For example:

- Name
- Address
- City
- State

To input the file format for fixed files, create a text document with the column names and the length of each column on its own line. For example:

- Name,25
- Address,40
- City,20
- State,2

Then input this file as the file format. The name of the text file will be the name of the file format.

Column length Mismatch between Fixed File and File Format

For Fixed Files, caution should be taken to ensure that the column length is in accordance with the File Format definition. Failure to do so will result in masking a column with the incorrect offset, which would have the unintended consequence of not masking what was intended.

Behavior when the number of fields in a delimited file's format and contents are mismatched

The behavior in this case is as follows:

1. If the total number of fields in the Delimited File is less than the total number of fields in the File Format, then after masking, delimiter will be added to match the total fields with File Format. See the below example, **Format:** One, Two, Three **Delimited File Data:** Test Data1, Test Data2 **Result after masking:** Test Data1, Test Data2, (One extra delimiter will be added to match with the File Format column length).
2. If the total number of fields in the Delimited File is greater than the total number of fields in the File Format, then after masking the extra fields in the Delimited File will be lost. See the below example, **Format:** One, Two, Three **Delimited File Data:** Test Data1, Test Data2, Test Data3, Test Data4 **Result after masking:** Test Data1, Test Data2, Test Data3

 **Multi-byte Characters** For Fixed Files, column length is determined by the number of characters rather than the number of bytes.

To import a new file format

1. Click **Import Format** at the upper right. The Import File Format window appears.
2. Select an **Import File Type**.

For a format type of copybook or XML

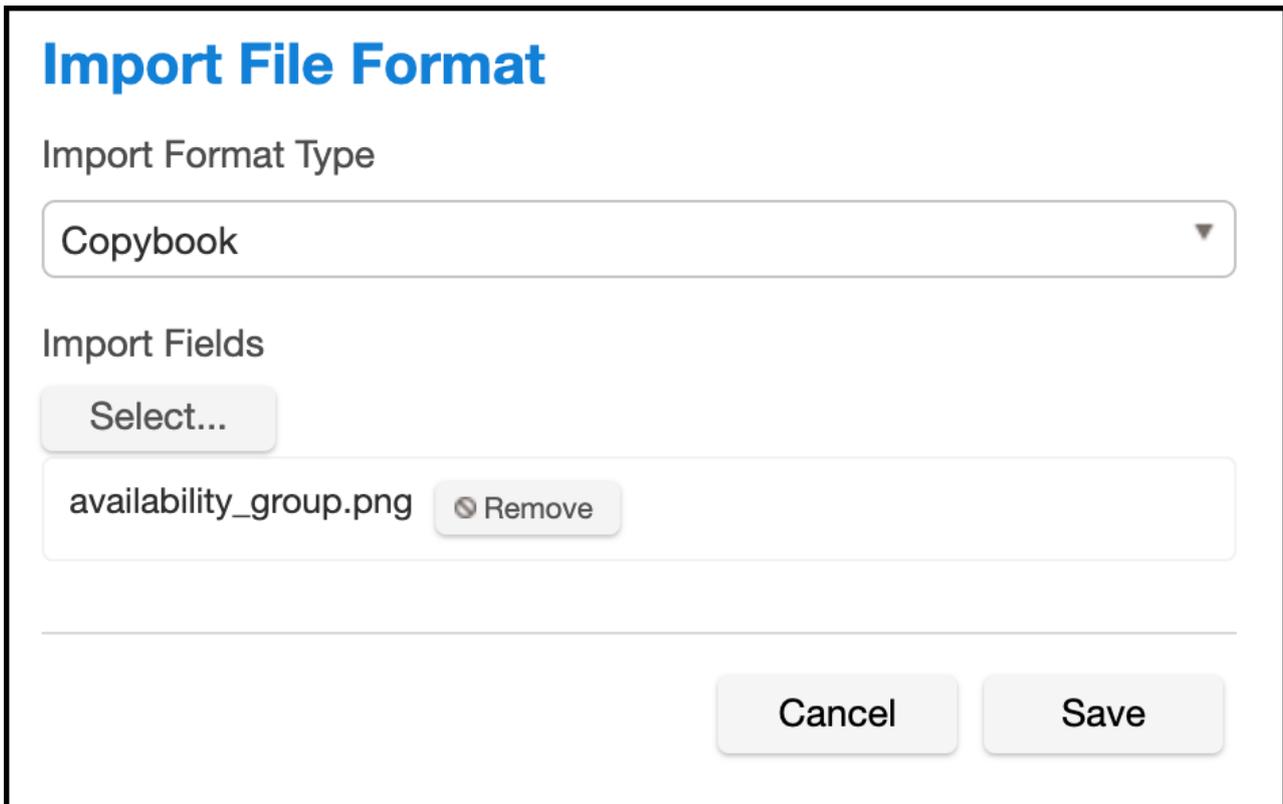
1. Select a **Connection Mode**.
2. Fill out the required fields of the selected **Connection Mode**. For Filesystem Mount Point connection mode, refer to the [Managing Remote Mounts](#) page to fill out the required fields.
3. Click **Browse**.
4. Click the **Select** button to the right of the desired import file format.
5. Enter a **Logical Name**.
6. Click **Submit**.

For a format type of delimited file, or fixed width file

1. Click **Select**.

2. Browse for the file from which to import fields.
Note: The contents of the imported file vary for Delimited, Fixed Width, Copybook (Mainframe), and XML types.
3. Click **Save**.
Note: The file must have NO header. Make sure there are no spaces or returns at the end of the last line in the file. To be masked, the field names must be in the same order as they are in the file.

Removing a selected file



Import File Format

Import Format Type

Copybook

Import Fields

Select...

availability_group.png Remove

Cancel Save

If you accidentally selected an incorrect file, simply click the Remove button to the right of the file and repeat the selection steps above.

Samples

The following is sample file content for Delimited file formats. With these formats, just the field name is provided. Notice there is no header and only a list of values.

```
First_Name  
Last_Name  
DOB  
SSN  
Address  
City  
State  
Zip_Code
```

The following is sample file content for Fixed Width format. In this format, the field name is followed by the length of the field, separated by a comma. Notice there is no header and only a list of values.

```
First_Name,20
Last_Name,30
DOB,10
SSN,11
Address,30
City,20
State,2
Zip_Code,10
```

To delete a file format

1. Click the **Delete** icon to the right of the File Format name.
2. File inventory is based on the file format. Therefore, if you make a change to a file inventory, that change applies to *all* files that use that format.
3. You can only add or delete a file format; you cannot edit one.

Assigning a file format to a files

Once you create a rule set with a file or set of files, you will need to assign those files to their appropriate file format. This is accomplished by editing the rule set. When you click on the edit button for the file a pop-up screen called edit file will appear with the file name. There will be a drop-down for the format so you can select the proper format for the file. If the file is a Mainframe data sets file with a copybook you will see a checkbox to signify if the file is variable length. For all other file types, select the end-of-record to let Delphix know whether the file is in Windows/DOS format (CR+LF) or Linux format (LF). If the file is a delimited file you will have a space to put in the delimiter. If there are multiple files in the ruleset you will have to edit each one individually and assign it to the appropriate file format.

Managing inventories

Managing inventories

An inventory describes all of the data present in a particular ruleset and defines the methods which will be used to secure it. Inventories typically include the table or file name, column/field name, the data classification, and the chosen algorithm.

The inventory screen

From anywhere within an environment, click the **Inventory** tab to see the Inventory Screen. This displays the inventory for the environment's rule sets.

Inventory settings

To specify your inventory settings:

1. On the left-hand side of the screen, select a **Rule Set** from the drop-down menu.
2. Below this, Contents lists all the tables or files defined for the ruleset.
3. Select a **table** or **file** for which you want to create or edit the inventory of sensitive data. The **Columns** or **Fields** for that specific table or file appear.
4. If a column is a primary key (PK), Foreign Key (FK), or index (IDX), an icon indicating this will appear to the Right of the column name. If there is a note for the column, a Note icon will appear. To read the note, click the icon.
5. If you selected a table, metadata for the column appears: **Data Type** and **Length** (in parentheses). This information is read-only.
6. Choose how you would like to view the inventory:
 - **All Fields** — Displays all columns in the table or all fields in the file (allowing you to mark new columns or fields to be masked).
 - **Masked Fields** — Filters the list to just those columns or fields that are already marked for masking.
 - **Auto** — The default value. The profiling job can determine or update the algorithm assigned to a column and whether to mask the column.
 - **User** — The user's choice overrides the profiling job. The user manually updates the algorithm assignment, mask/unmask option of the column. The Profiler will ignore the column, so it will not be updated as part of the Profiling job.

DELPHIX MASKING Job Wizard admin

[Environments](#) [Monitor](#) [Settings](#) [Admin](#) [Audit](#)

[Overview](#) [Connector](#) [Rule Set](#) [Inventory](#)

Home > Environments > test1 > Inventory > test

[Import](#) [Export](#)

Filter By: [All Fields](#) [Masked Fields](#) [Auto](#) [User](#)

Column	Data Type	Algorithm	Edit
fname	varchar (50)		
idd (PK ID IX)	int (0)		
lname	varchar (50)		

Select Rule Set

test

Filter Contents

Search By Name

Search Alphabetically

Show Masked Tables

Contents

- DBVERIFICATION_TA...
- Foo
- Foo1**
- foo1010101
- foo111122
- foo121333
- Foo2
- foo_test

[Environments](#) | [Monitor](#) | [Settings](#) | [Admin](#) | [Audit](#) DELPHIX

Assigning algorithms

To set criteria for sensitive columns or fields:

1. Click the edit icon to the right of a column or field name.
2. From the **Domain** drop-down list, select the appropriate sensitive data element type.
3. The Delphix Masking Engine defaults to a **Masking Algorithm** as specified in the Settings screen. If necessary, you can override the default algorithm.
 - To select a different masking algorithm, choose one from the **Algorithm** drop-down list. For detailed descriptions of these algorithms, see [Out Of The Box Algorithm Frameworks](#).
4. Select an **ID Method**:
 - **Auto** — The default value. The profiling job can determine or update whether to mask a column.
 - **User** — The user decides whether to mask/unmask a column. The user's choice overrides the profiling job. (The user masking is done after the profiling job is finished.)
5. You can add/remove notes in the **Notes** text field.
6. When you are finished, click **Save**. You must click Save for any edits to take effect.

- ☐ If you select a DATESHIFT algorithm and you are not masking a datetime or timestamp column, you must specify a **Date Format**. (This field only appears if you select a DATESHIFT algorithm from the Masking Algorithm dropdown.) For a list of acceptable formats, click the **Help** link for Date Format. The default format is yyyy-MM-dd.

Managing a file inventory

Defining fields

- ☐ You must select a delimited or fixed-width file connector from the **Select Rule Set** drop-down list on the left navigation pane, not a database.

To create new fields:

1. From an Environment's Inventory tab, click the **Define fields** to the far right. The Edit Fields window appears.
2. Edit the fields as described in **Setting Field Criteria for a File**.
3. When you are finished, click **New** to create a new field, or click **Save** to update an existing field.

Managing a mainframe inventory

Redefine conditions

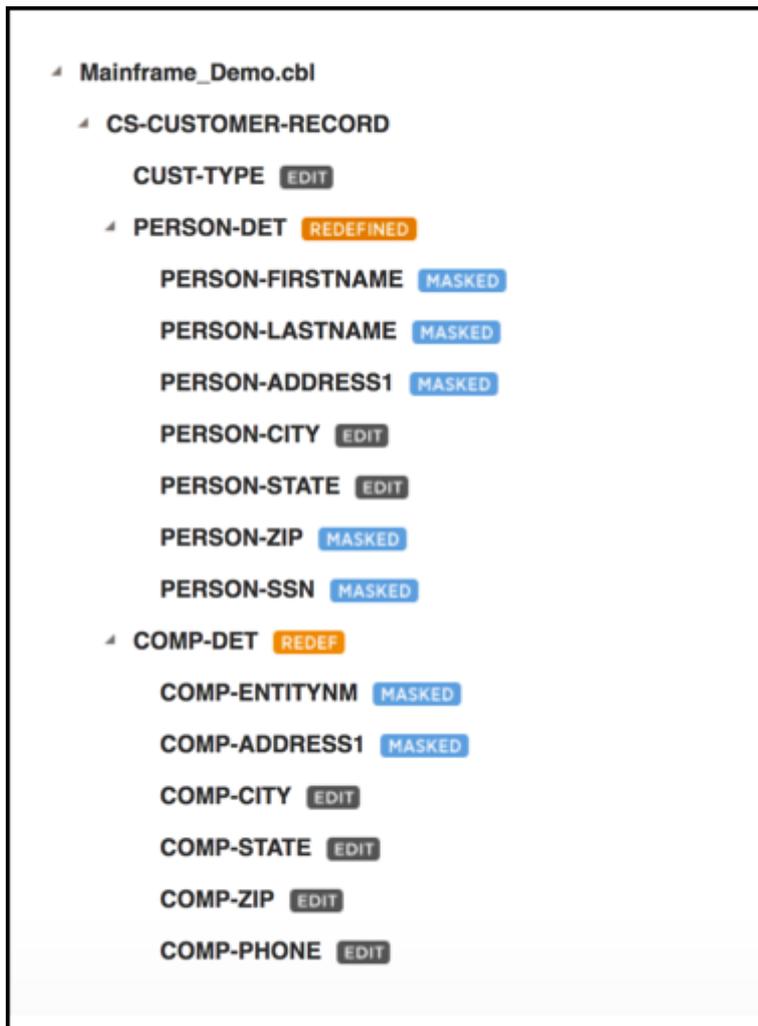
For Mainframe data sets, the inventory also allows for the entry of Redefine Conditions, which are used to handle any occurrences of COBOL's REDEFINES construct that might appear in the Copybook. In COBOL, the REDEFINES keyword allows an area of a record to be interpreted in multiple different ways. In the example below, for instance, each record can hold either the details of a person (PERSON-DET) or the details of a company (COMP-DET).

```

01 CS-CUSTOMER-RECORD.
  05 CUST-TYPE                PIC X(1) .
  05 PERSON-DET.
    10 PERSON-FIRSTNAME      PIC X(20) .
    10 PERSON-LASTNAME      PIC X(40) .
    10 PERSON-ADDRESS1      PIC X(50) .
    10 PERSON-CITY          PIC X(20) .
    10 PERSON-STATE         PIC X(5) .
    10 PERSON-ZIP           PIC X(10) .
    10 PERSON-SSN           PIC S9(9) COMP-3.
  05 COMP-DET                 REDEFINES PERSON-DET.
    10 COMP-ENTITYNM        PIC X(53) .
    10 COMP-ADDRESS1       PIC X(50) .
    10 COMP-CITY           PIC X(20) .
    10 COMP-STATE         PIC X(5) .
    10 COMP-ZIP           PIC X(10) .
    10 COMP-PHONE         PIC X(12) .

```

Depending on which group is present, different masking algorithms may need to be applied. Below is the inventory corresponding to this copybook, which allows algorithms to be selected separately for each group.



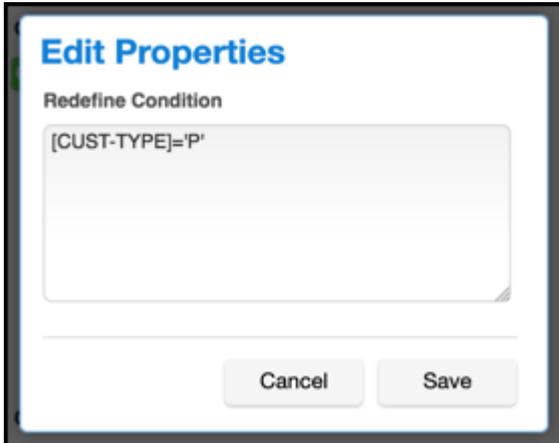
In order to do any masking however, the Masking Engine must be able to determine, for each record, which fields should be read, so that the correct algorithms can be applied. In order to do this, the masking engine uses Redefine Conditions, which are specified in the inventory. Redefine Conditions are boolean expressions that can reference any fields in the record when they are evaluated.

In the example copybook above, the field CUST-TYPE is used to indicate which group is present. If CUST-TYPE holds a 'P', a PERSON-DET group is present, and if it holds a 'C', COMP-DET is present. This can be expressed in the inventory by specifying a Redefine Condition with the value [CUST-TYPE]='P'. This expression indicates that, for each record read from the source file during the masking job, the value of the field CUST-TYPE should be read and compared against the string 'P'. If it is equal, the Masking Engine will read from the record the fields subordinate to PERSON-DET, and will apply any masking algorithms specified on those fields. Similarly, a Redefine Condition with the value [CUST-TYPE]='C' should be applied to the COMP-DET field. Exactly one of the conditions should evaluate to 'true' for each group of redefined fields. For example, a copybook might have fields A, B REDEFINES A, and C REDEFINES A. Of the Redefine Conditions attached to A, B, and C, one and only one should evaluate to true for each record.

Entering a redefine condition

1. Click on the orange **REDEFINED** or **REDEF** button next to the redefined or redefining field
2. Enter a condition in the dialog box which appears. This is the expression, which, when it evaluates to true, causes the subordinate fields to be read and, if they have algorithms assigned, masked.

3. Click **Submit**.



Format of redefine conditions

Redefine Conditions allow fields to be compared against either number or string literals. Square brackets enclosing a field name indicate a variable, which takes on the value of the named field:

```
[Field1] = 'An example String'
```

String literals can be enclosed in either single or double quotes. For fields that are numeric (e.g. PIC S99V9), the operators <, <=, >, and >= can be used in addition to the =operator, e.g.

```
[Field2] <= -10.5
```

Also, conditions can be joined using AND, OR, and NOT to form more complex conditions:

```
([Field3] > 2.5 AND [Field3] < 10) OR NOT [FIELD4] = 'Z'
```

Importing and exporting an inventory

To export an inventory:

1. Click the **Export** icon at the upper right. The Export Inventory pop-up appears with the name of the currently selected Rule Set as the Inventory Name and a corresponding .csv **File Name**.
2. Click **Save**.

A status pop-up appears. When the export operation is complete, you can click on the **Download file** name to access the inventory file

To import an inventory:

1. In the upper right-hand corner, click the **Import** icon. The Import Inventory pop-up appears.
2. Click **Select** to browse for the name of a comma-separated (.csv) file.
3. Click **Save**.

The inventory you imported appears in the **Rule Set** list for this environment.



- You can only import one ruleset at a time.
- The format of an imported .csv file must exactly match the format of the exported inventory. If you plan to import an inventory, you should export it first and then update the exported file as needed before importing it.
- After importing the inventory to the 10.0.0.0 version Compliance engine from older versions, ruleset refresh is mandatory when the inventory has any document store type assignments or the user needs to perform document store type masking on the columns from the imported inventory.

Document store-type masking

This feature provides the ability to mask structured documents that are stored in database columns. This is done by marking a column as **Structured** and assigning a respective **Document Store Type** and **File Format** to it.

With the release of version 10.0.0.0 of the Continuous Compliance engine, the document store type masking will support automatic datatype identification. This will be done by using the [JDBC SQL Type](#) associated with columns. String and BLOB types will be supported for document store type masking.



With version 10.0.0.0 release

- In the case of existing rulesets, a ruleset refresh is mandatory before using Document Store Type masking.
- Masking jobs having rulesets with document store type assignments will need mandatory ruleset refresh. Without ruleset refresh job will not be allowed to run.
- Masking jobs having rulesets without document store type assignments will not need ruleset refresh.
- Ruleset refresh is not required for newly created rulesets.



- The column type should be from one of the following JDBC SQL Types: CHAR, NCHAR, VARCHAR, NVARCHAR, CLOB, NCLOB, LONGVARCHAR, LONGNVARCHAR, BLOB, SQLXML
- BLOB type will not be supported for MySQL databases.
- SQLXML type will be only supported for Oracle databases.
- The file format must be either XML or JSON

Columns with a supported data type have a setting called **Data Model**, which can be set to either *Plain* or *Structured* values.

As shown in the image below, columns with *Plain* selected as the Data Model can be masked as a single value by assigning a **Domain** and **Algorithm**.

Edit Properties

Column Name	Notes
<input type="text" value="xml_clob"/>	<div style="border: 1px solid #ccc; height: 80px;"></div>
ID Method	
<input type="text" value="Auto"/>	
Data Model Learn More	
<input type="text" value="Plain"/>	
Domain	Algorithm
<input type="text" value="NULL_SL"/>	<input type="text" value="NULL SL"/>

When the *Structured* value is selected for the Data Model, a **Document Store Type** and **File Format** can be assigned as shown in the image below.

Edit Properties

Column Name

ID Method

Data Model [Learn More](#)

Document Store Type

Notes

File Format

The image below shows the **Inventory** screen for a rule set with a structured column. To quickly access an assigned File Format from this screen (books.xml in this example), click on the file format's name in the **File Format** panel in the lower left.

Select Rule Set

Filter Contents

Contents

test

File Format

Column	Data Type	Algorithm	File Format	Edit
id (PKIX)	serial (10)			✎
json_clob	text (2147483647)			✎
xml_clob	text (2147483647)		books.xml	✎

Multi-column algorithm support

With the release of version 10.0.0.0, [Multi-column algorithms](#) will be supported for JSON and XML document store type masking with limited buffer-data size.

- i** Buffer size (in bytes) will be using calculated using the below formula:
 $((\text{Max_memory_of_Job}/\text{No_of_streams_for_job}) * \text{CharStreamingBufferLimitRate})/100$
- The default values will be used when the maximum memory and number of the stream for the job are not defined.
 - Buffer-data size is configurable via the application setting "**CharStreamingBufferLimitRate**" under Mask group settings. For adjusting "**CharStreamingBufferLimitRate**" refer to [Masking API Client](#).

The fields having multi-column assignments should not exceed the limit of buffer data size. In case of exceeding the limit of buffer data size, the job will fail. Users can configure buffer size by adjusting "**CharStreamingBufferLimitRate**" to avoid exceeding the buffer data size issue.

JSON file format

For details on Multi-column algorithm support with JSON file format refer to [JSON File Masking](#)

XML file format

In the case of XML document store type masking, Multi-column algorithm assignment to XML elements will not be validated at the time of assignment. Because in XML it is difficult to find out if an element is a type of an array or a single element until the whole data is read. Here, the masking job will fail immediately when any of the invalid multi-column assignments are found while running the job. Make sure the algorithm assignment should follow the below rules.

- i**
- Multi-column algorithm for XML file masking is not supported.
 - Multi-column algorithm assignment to XML attributes is not supported.
 - Multi-column algorithm is not supported for XML elements where,
 - The element is a type of array.
 - Elements are part of different arrays.
 - Elements are on different levels having one or more elements of type array.

Below is a sample XML file format with valid and invalid multi-column assignment examples.

XML Content	XML elements having Multi-column assignment	Invalid/Valid
<pre> <data> <founders> <name>Whooley</name> <name>Boddy</name> </founders> <hospital id="101"> <name>Hospital_name</name> <registered_on>2001-01-01</registered_on> <inaugurated_on>2002-01-01</inaugurated_on> <found_date>1905-12-10</found_date> <patient> <doa>2001-07-31</doa> <dob>1907-08-01</dob> <dol>2005-04-12</dol> <name>Patient_Name1</name> </patient> <patient> <doa>2002-09-31</doa> <dob>1908-01-01</dob> <dol>2007-04-11</dol> <name>Patient_Name2</name> </patient> </hospital> <registration_id>932884</registration_id> <doctor_details> <doj>2001-07-31</doj> <dol>2004-01-02</dol> <doctor_name>Doctor_Name1</doctor_name> </doctor_details> <doctor_details> <doj>2005-07-07</doj> <dol>2007-01-03</dol> <doctor_name>Doctor_Name2</doctor_name> </doctor_details> </data> </pre>	<pre> /data/founders/name /data/registraion_id </pre>	<p>Invalid. Multi-column algorithm assignment is not allowed as element 'name' is an array</p>
	<pre> /data/hospital/patient/name /data/doctor_details/doctor_name </pre>	<p>Invalid. Multi-column algorithm assignment is not allowed for elements that are part of different arrays. Here 'name' belongs to an array 'patient' and 'doctor_name' belongs to an array 'doctor_details'</p>
	<pre> /data/hospital/found_date /data/hospital/patient/doa </pre>	<p>Invalid. Multi-column algorithm assignment is not allowed for these as one of the elements i.e. 'doa' belongs to an array and elements are on different levels.</p>
	<pre> /data/hospital/patient/doa /data/hospital/patient/dol </pre>	<p>Valid</p>
	<pre> /data/hospital/registered_on /data/hospital/inaugured_on </pre>	<p>Valid</p>
	<pre> /data/doctor_details/doj /data/doctor_details/dol </pre>	<p>Valid</p>
	<pre> /data/hospital/name /data/registration_id </pre>	<p>Valid</p>
	<pre> /data/hospital/@id /data/hospital/name </pre>	<p>Invalid. Multi-column algorithm assignment is not allowed for XML attributes.</p>

Managing record types

Overview

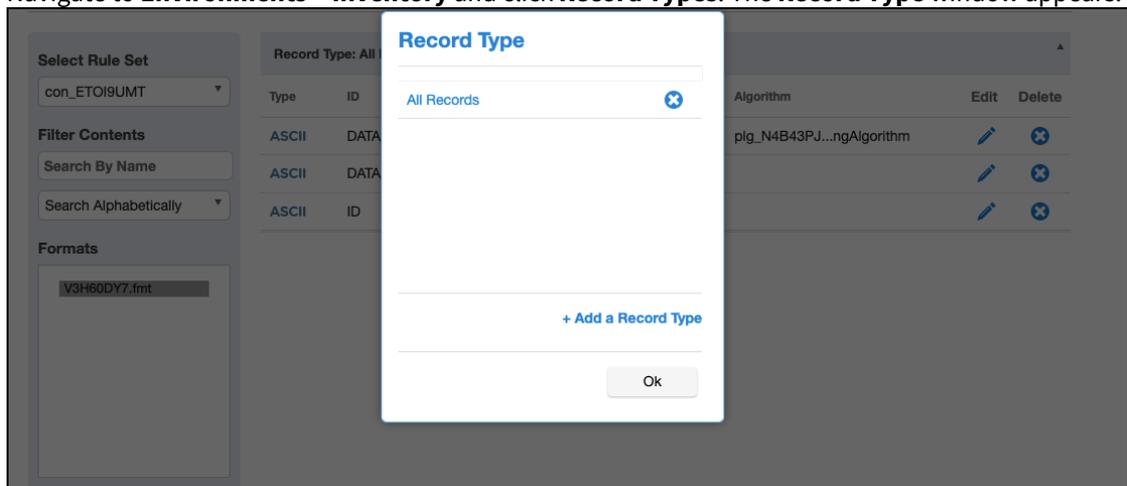
You can use record types to perform conditional masking of the file records. If a file has a different set of records spread across multiple rows, then the masking engine should be able to understand all the unique records. For example, a file has the following record in the first three columns of each row: first name, last name, and age. But the last column of each row has a unique record like IP address, ethernet address, etc. Then you must create a new record type for every unique record present in the file and assign a specific file format to all the record types.

F You must select a rule set that was created using a file connector from the **Select Rule Set** dropdown list on the left navigation pane. Record types are applicable only for delimited and fixed-width file type connectors.

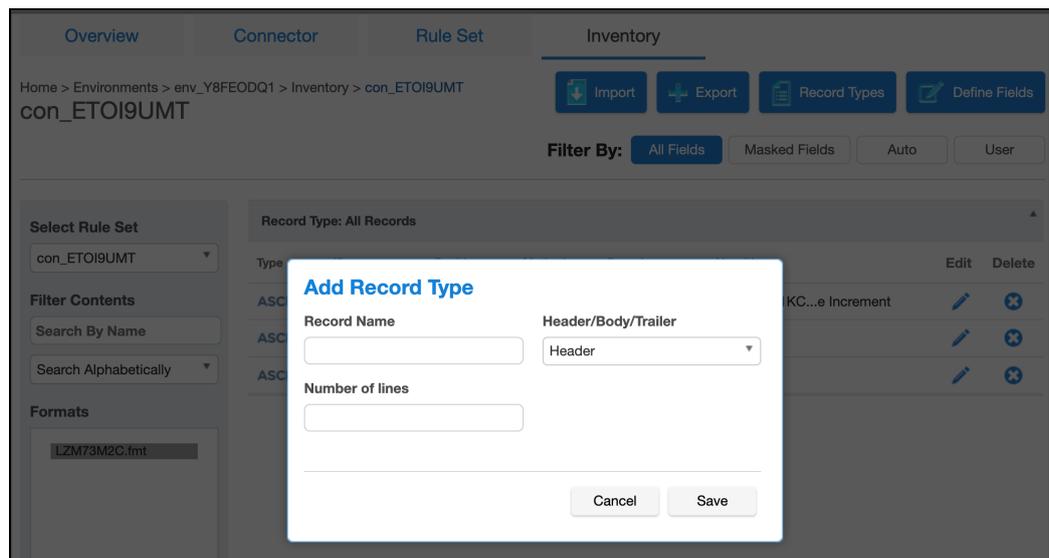
Adding record types

Perform the following steps to add a record type:

1. Create a record type for each distinct type of record (for each distinct list of fields) and assign each a file format.
 - a. Navigate to **Environments > Inventory** and click **Record Types**. The **Record Type** window appears.



- b. Click **+Add a Record Type** at the bottom of the window. The **Add Record Type** window appears.
- c. In the **Add Record Type** window, enter values for the following fields:
 - i. **Record Name** — A free-form name for this record type.
 - ii. **Header/Body/Trailer** — Select one of the following: Header, Body, or Trailer. Delphix allows the masking of multiple types of body records. If the file has header or trailer records, you will need to create record types for them. The **Header** or **Trailer** record type is used to specify a number of records that are not masked at the beginning and end of a file. If you selected **Header** or **Trailer**, then enter the **Record Name** and the **Number of lines** for the header/trailer.



If you selected **Body**, then you must do the following:

1. **# of Identifier fields** - Specify the number of identifier fields.
2. **Import Fields** - Click on the **Select** button to browse for the file from which to import fields.

Note

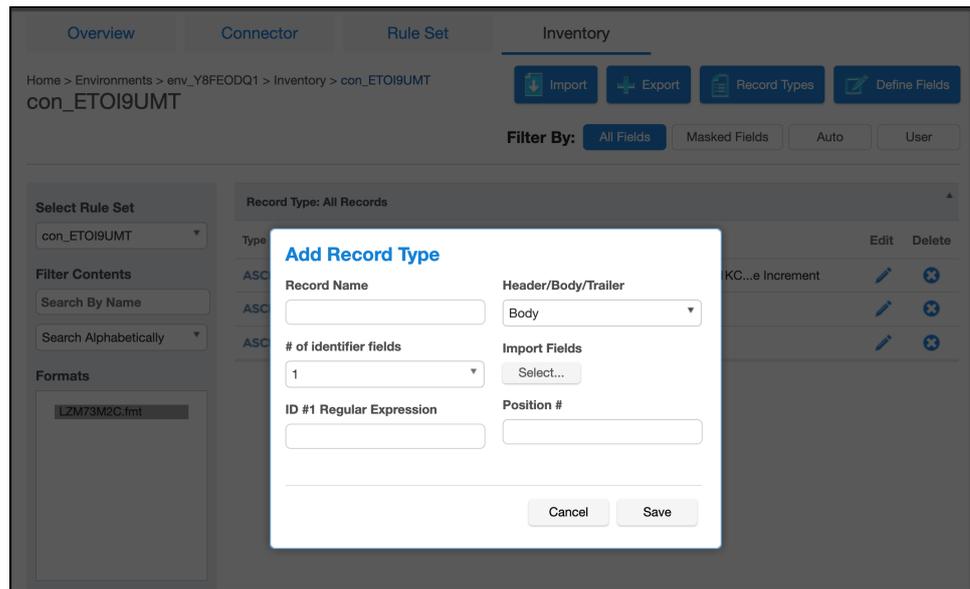
The contents of the imported file vary for Delimited, Fixed Width, Copybook (Mainframe), and XML types.

3. **ID # 1 Regular Expression**— (optional) Specify the value of the record type code or another identifier that allows Delphix to identify records that qualify as this record type. A record type applies if its regular expression matches its specified identifier fields.

Info

This value is a regular expression that the masking engine uses to match the specified field to determine whether the record is of this type. For example, the expression "C_{[A-Z]}{2}" can be used to match C_IP for the type C records (as described in step 1 example).

4. **Position #** — (optional) Specify the field number (for delimited files) or the character position number (for fixed files) of the beginning of the Record Type Identifier within the data record.



- d. Click **Save** when you are finished.
2. Modify inventory for each record type by clicking the Edit icon next to the ID of each record.
3. Create and run a masking job. For more information, see [Creating Masking Jobs](#).

Masking whole file

You can now configure the masking engine to mask the complete file (for example, JSON or Parquet) and pass the content of that file as a single input to an algorithm.

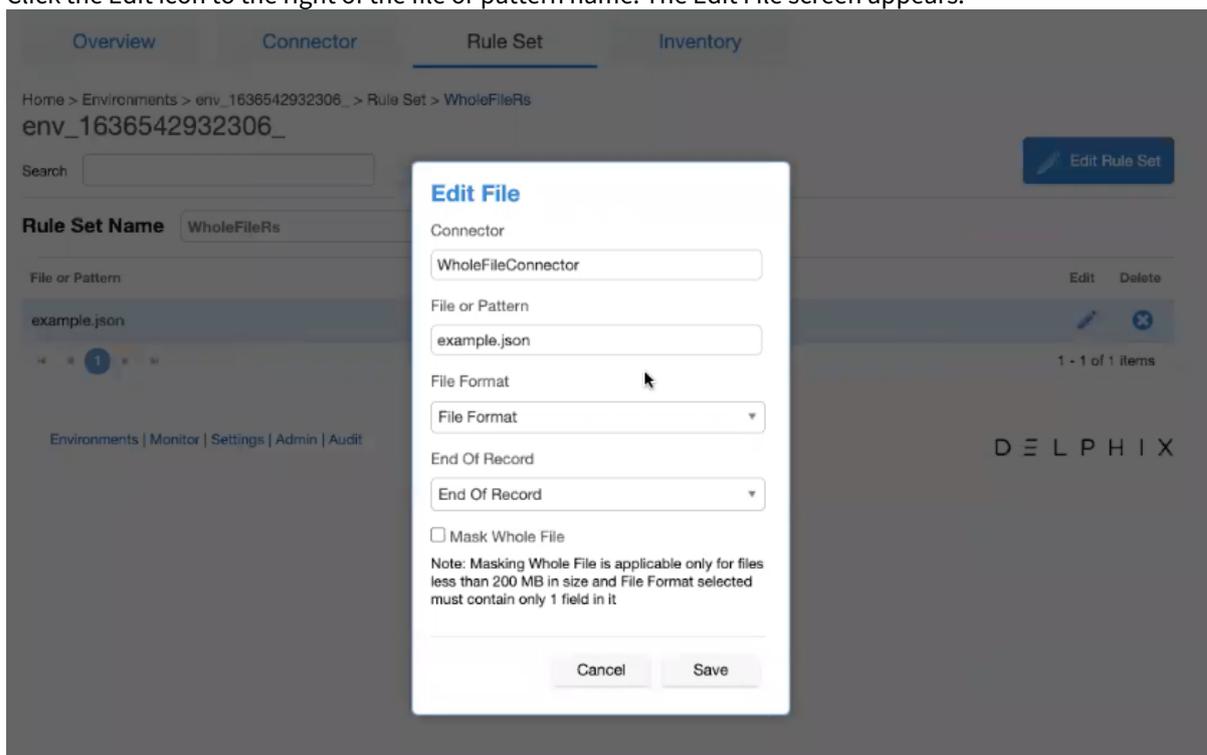
Pre-requisite

- You must create a fixed-width file connector. For more information on creating connectors, see [Managing connectors](#).
- You must create a fixed-width file format that has only one field defined in it. For more information on creating file formats, see [Managing file formats](#).

Masking a whole file

Perform the following procedure to mask a whole file.

1. Navigate to **Environments > Ruleset**.
2. On the **Rule Set** screen, click the Edit icon to the right of the fixed-width connector rule set. Alternatively, click on the fixed-width connector name. The rule set screen displays all the files in the directory that are associated with the respective connector.
3. Click the Edit icon to the right of the file or pattern name. The Edit File screen appears.

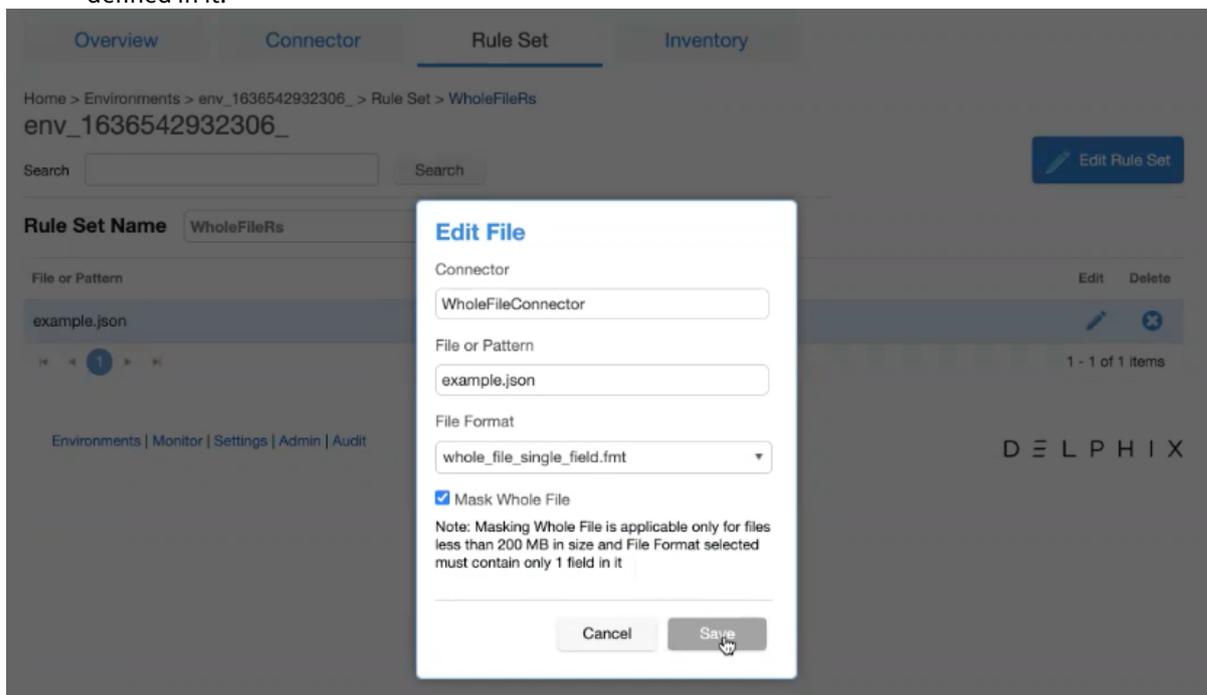


4. From the **File Format** drop-down list, select a file format that has only one field defined in it. Selecting any other file format will result in an error.
5. Select the **Mask whole file** checkbox to enable whole file masking. Selecting this option results in the disappearance of other options (End of Record, Delimited, Enclosure, and Escape Character for Enclosure). These configurations are no more required as the masking engine will now read the whole file and send it to the algorithm.

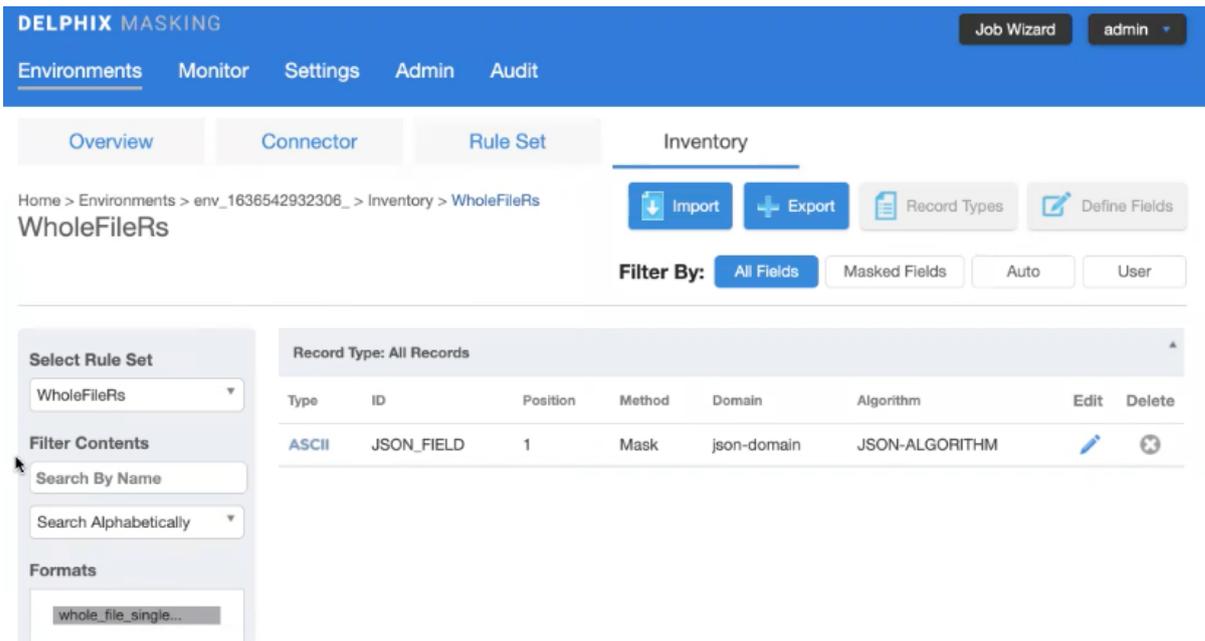
Note

Masking the whole file is applicable only for:

- Files that are less than 200 MB in size. However, you can modify this limit via API by configuring **Whole File Masking Max File Size In MB** key in the Application Settings.
- The file format that has only one field defined in it. The masking whole file is applicable only for: Files that are less than 200 MB in size. However, you can modify this limit via API by configuring **Whole File Masking Max File Size In MB** key in the Application Settings. File format that has only one field defined in it.

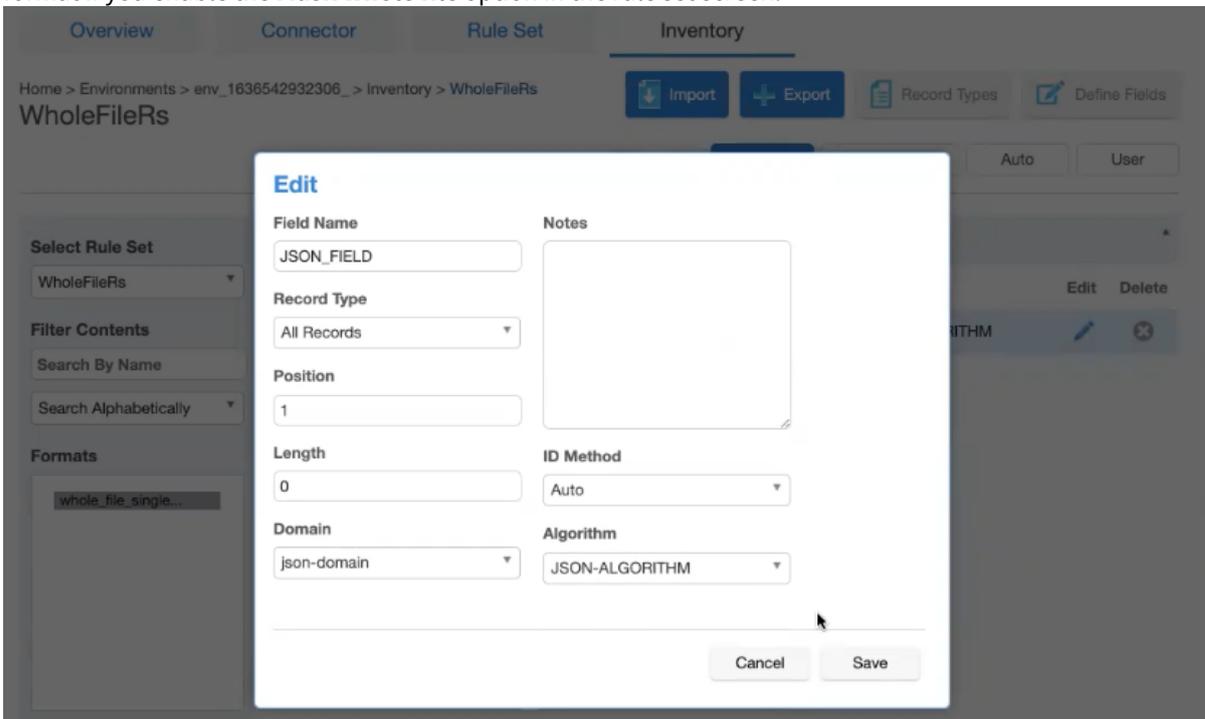


6. Click **Save**.
7. Navigate to **Environments > Inventory**. The **Record Types**, **Define Fields**, and **Delete** options are greyed-out when you select a file format that is used with any fixed-width file having **Mask whole file** option enabled.



8. Click the Edit icon to the right of the record type.
9. From the **Algorithm** drop-down list, select the matching extended algorithms that must be applied to the file.

Note: You can not modify/update the length and position for the single field defined in the respective file format if you enable the **Mask whole file** option in the rule set screen.



10. Click **Save**.

JSON file masking

Introduction

This feature offers standard functionalities for masking JSON files. Users will now be able to configure and run Continuous Compliance jobs specific to JSON files, assigning algorithms to any field of a JSON file using their respective JSON paths. This feature overcomes the shortfalls of the existing algorithm-based workaround by providing users with a simplified way to assign Continuous Compliance algorithms. This feature also supports masking JSON files of large sizes.

These features are not yet supported:

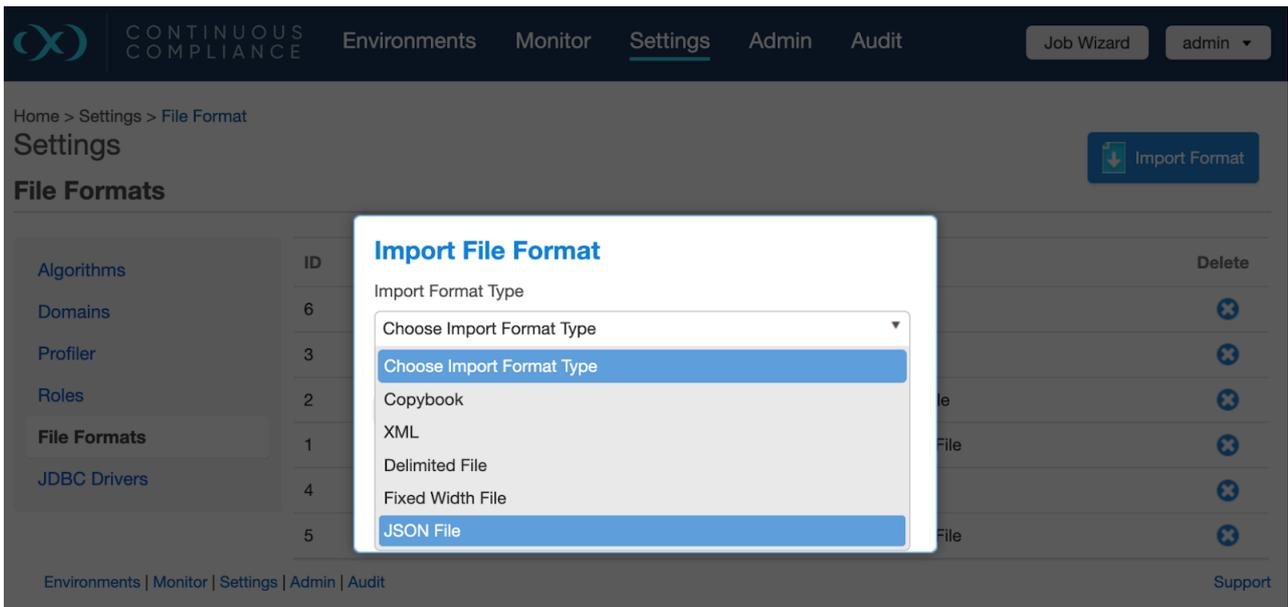
- Profiling Job for JSON File Rulesets

API changes

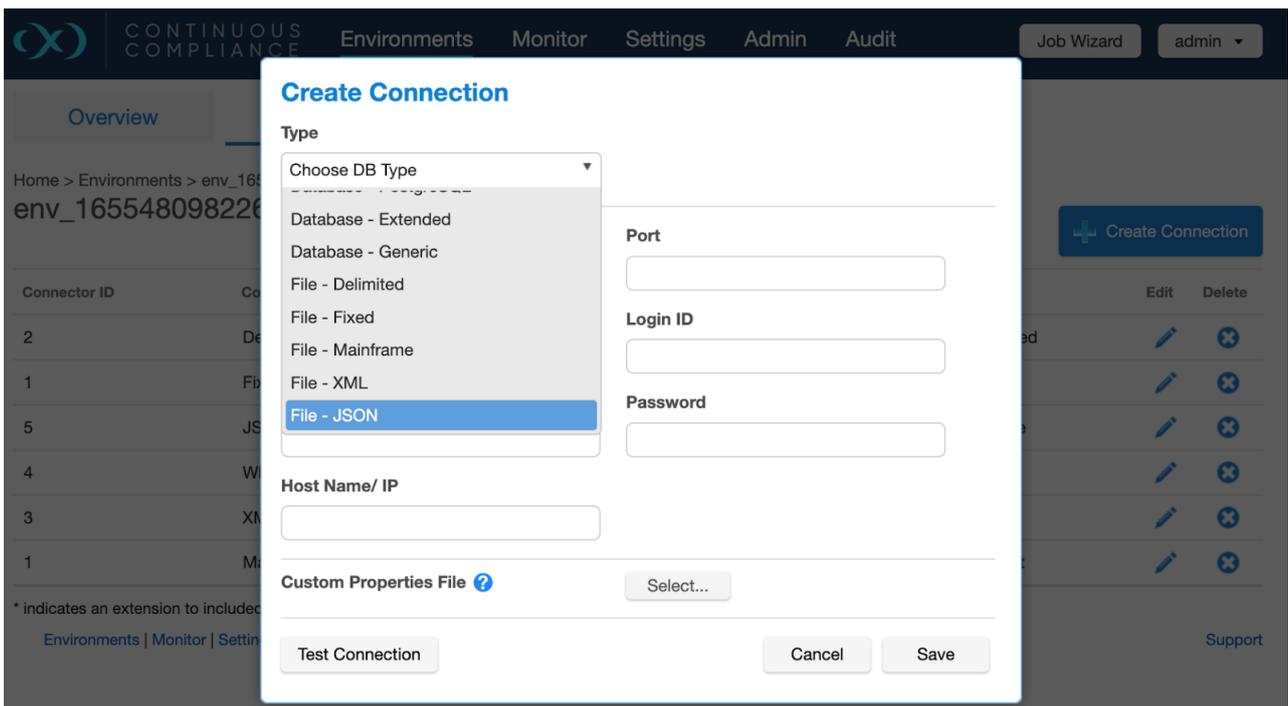
API	Change Description
POST /file-formats	Added support to upload a Json file to create JSON File Format.
PUT /file-formats/	Added validations to stop creating headers and footers for JSON File Formats.
POST /file-connectors	Added support to create a new file connector of type <code>File - JSON</code> .
POST /file-field-metadata	Added support to create a new JSON File field, specifying its JSON path identifier and assigning algorithms to it.
PUT /file-field-metadata	Added support to update JSON File field to assign or unassign algorithms to it.

GUI changes

In the Continuous Compliance interface, navigate to **Settings > File Format**. Import the .json file to create JSON File Formats. Use the file you want to mask as the format.



In the Create Connection screen, choose **File - JSON** from the Type dropdown and configure the appropriate details.



Below is a JSON file example:

```
[
  {
    "id":1,
    "first_name":"abc",
    "last_name":"xyz",
    "email":"abc.xyz@gmail.com",
    "gender":"Female",
    "ip_address":"26.58.193.2",
    "dob":"2002-07-19"
  }
]
```

The **Inventory** tab for JSON File Formats is used to configure algorithms to JSON Paths and to add new JSON paths using Define Fields button.

Home > Environments > env_1654891654749_ > Inventory > JSONFileRs

 Define Fields

JSONFileRs

Filter By: All Fields Masked Fields Auto User

Select Rule Set

JSONFileRs

Filter Contents

Search By Name

Search Alphabetically

Formats

alltype.json

File

huge_alltype.json

Record Type: All Records

JSON Path	Domain	Algorithm	Edit	Delete
\$[*]['str_arr']	FIRST_NAME	FIRST NAME SL		
\$[*]['str_arr']				
\$[*]['price']				
\$[*]['name']	FIRST_NAME	FIRST NAME SL		
\$[*]['long']				
\$[*]['int']				
\$[*]['dob']				
\$[*]['contact']['phone']				
\$[*]['contact']['email']				
\$[*]['contact']['address']['state']				
\$[*]['contact']['address']['pincode']				
\$[*]['contact']['address']['country']				

Navigate to **Monitor > Processing** to access the Job Process Monitoring page. This page shows data in byte format for JSON file masking.


CONTINUOUS COMPLIANCE
Environments
Monitor
Settings
Admin
Audit
Job Wizard
admin ▾

Home > Monitor > Processing
Monitor

FILE
RUNNING
1 Jobs Running

JSON File Maski...

Job Type Mask Environment env_16548916547.. Job ID 4 Execution ID 51 CM Connection jsonFile Source / Target - / JSONFileConnect..	<ul style="list-style-type: none"> ✓ Init Execution ✓ Collecting Job Configurations ✓ Preparing Execution ⌚ Execute Pre Execution Custom Driver Task ⌚ Start Execution ⌚ Pre SQL Script ⌚ Post SQL Script ⌚ Execute Post Execution Custom Driver Task 9 Collecting Job Information 10 Execution Finished 	Start Time 04:37:35 Previous Run Time 00:14:21 Total # of Files 2 Files Masked 1 Files with Nonconforming Data 0 Files to be Masked 1 Estimated run time (HH:mm:ss) 00:14:00 Bytes Remaining 1.60 GB Bytes Processed 382.42 MB Fields with Nonconforming Data 0 Streams 1
--	--	--

Completed
Processing
Waiting

Processing 0 In Queue

ID	Name	Progress	ETA (HH:mm:ss)	Bytes Per Min	Bytes Processed	Bytes Remaining
59	huge_alltype.json	<div style="width: 18%; background-color: #c6e0b4;">18%</div>	00:13:23	122.46 MB	382.42 MB	1.60 GB

[Environments](#) | [Monitor](#) | [Settings](#) | [Admin](#) | [Audit](#)

[Support](#)

Constructing a JSON file path

A JsonPath expression begins with the dollar sign (\$) character, which refers to the root element. The dollar sign is followed by a sequence of child elements, which are separated by the square brackets (['']) containing the name of each JSON field. If the field is inside an array, a star character is used to represent all elements of the array ([*]).

Json Content	Bracket Separated
<pre>{"firstName" : "xyz"}</pre>	<pre>\$['firstName']</pre>
<pre>[{ "firstName" : "xyz" }]</pre>	<pre>\$[*] \$[*]['firstName']</pre>
<pre>{ "employee": { "contacts" : ["987654321", "080-23456789"] } }</pre>	<pre>\$['employee'] \$['employee']['contacts'] \$['employee']['contacts'][*]</pre>

Multi-column algorithm support

Starting with version 10.0.0.0, JSON file masking with limited buffer-data size will support [multi-column algorithms](#). This enables the use of multiple algorithms to mask data in JSON files, even if the file is large.

= Buffer size (in bytes) will be calculated using the formula below:

$$((\text{Max_memory_of_Job}/\text{No_of_streams_for_job}) * \text{CharStreamingBufferLimitRate}) / 100$$

- The default values will be used when the maximum memory and number of the stream for the job are not defined.
- Buffer-data size is configurable via the application setting "CharStreamingBufferLimitRate", under **Mask group** settings. To adjust " CharStreamingBufferLimitRate", refer to [Masking API Client](#).

The fields having multi-column assignments should not exceed the limit of buffer data size. In case of exceeding the limit of buffer data size, the job will fail. Users can configure buffer size by adjusting "CharStreamingBufferLimitRate" to avoid exceeding the buffer data size limit issue.

- =**
- Multi-column algorithm is supported for JSON File and JSON [Document store type](#) masking.
 - Multi-column algorithm is not supported for JSON fields where,
 - JSON field is an array.
 - JSON fields are part of different arrays.
 - JSON fields are on different levels having one or more fields from JSON arrays.

Multi-column algorithm assignment for JSON fields will be validated at the time of assignment. If any of the above combinations is found while assigning a multi-column algorithm, that assignment will not be allowed.

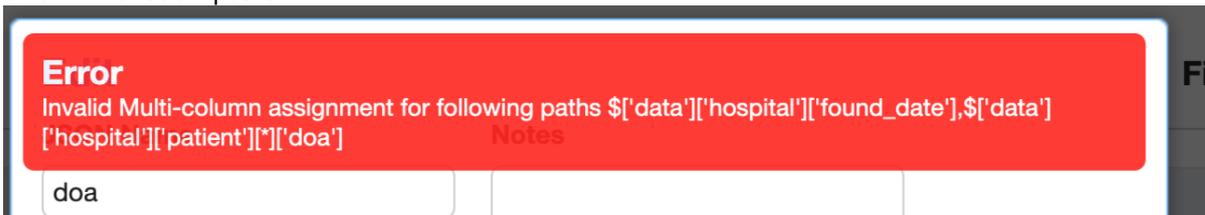
Below is a sample JSON file format with valid and invalid multi-column assignment examples.

JSON Content	JSON Fields having Multi-column assignment	Invalid/Valid
<pre> { "data": { "founders": { "name": ["Founder_Name1", "Founder_Name2"], "hospital": { "name": "Hospital_Name", "registered_on": "2001-01-01", "inaugurated_on": "2002-01-01", "found_date": "1905-12-10", "patient": [{ "doa": "2001-07-31", "dob": "1907-08-01", "dol": "2005-04-12", "name": "Patient_Name1", "test_report": [{ "test_id": 1, "test_name": "TSH1", "dot": "2001-08-02", "result": "positive" }] }] }, "registration_id": 932884, "doctor_details": [{ "doj": "2001-07-31", "dol": "2004-01-02", "doctor_name": "Doctor_Name1" }] } } } </pre>	$\$['data']['hospital']['name'][*]$ $\$['data']['registration_id']$	Invalid. Multi-column algorithm assignment is not allowed as JSON field 'name' is an array
	$\$['data']['hospital']['patient'][*]['name']$ $\$['data']['doctor_details'][*]['doctor_name']$	Invalid. Multi-column algorithm assignment is not allowed for JSON fields that are part of different JSON arrays. Here 'name' belongs to an array 'patient' and 'doctor_name' belongs to an array 'doctor_details'
	$\$['data']['hospital']['found_date']$ $\$['data']['hospital']['patient'][*]['doa']$	Invalid. Multi-column algorithm assignment is not allowed for these fields as one of the fields i.e. 'doa' belongs to an array and fields are on different levels.
	$\$['data']['hospital']['patient'][*]['doa']$ $\$['data']['hospital']['patient'][*]['dol']$	Valid
	$\$['data']['hospital']['registered_on']$ $\$['data']['hospital']['inaugurated_on']$	Valid
	$\$['data']['doctor_details'][*]['doj']$ $\$['data']['doctor_details'][*]['dol']$	Valid
	$\$['data']['hospital']['name']$ $\$['data']['registration_id']$	Valid
	$\$['data']['hospital']['patient'][*]['dol']$ & $\$['data']['hospital']['patient'][*]['test_report'][*]['dot']$	Invalid. Multi-column algorithm assignment is not allowed for these fields as they are different levels having fields belonging to different arrays. 'dol' belongs to 'patient' while 'dot' belongs to 'test_report'

Error management

Multi-column algorithm is supported with some limitations so there are different errors possible.

- In case of assigning a multi-column algorithm to JSON fields with an invalid combination error will be thrown with JSON paths.



- Multi-column algorithm assignment and any type of algorithm assignment to JSON multi-dimensional array field i.e. $\$['sample'][*][*]$ type of paths is not allowed when they are under the same parent level. While assigning inventory to this type of combination error will be thrown.
 - For example, using the below JSON, if the multi-column algorithm is assigned to fields,
 - $\$['data']['date1']$
 - $\$['data']['date2']$ then algorithm assignment to $\$['data']['Sample_2'][*]$ will not be allowed, as they are under same parent $\$['data']$.

```
{
  "data": {
    "Sample_1": [
      "ABCDEF",
      "ABCDEF"
    ],
    "Sample_2": [
      [
        "ABCDEF",
        "ABCDEF"
      ],
      [
        "ABCDEF",
        "ABCDEF"
      ]
    ],
    "date1": "2021-02-08",
    "date2": "2022-02-09"
  }
}
```

Error

Algorithm assignment to multi-dimensional field is not allowed, when multi-column algorithm is assigned to fields present within same parent level. For more information, see <https://www.delphix.com/masking-help/document-store-help>

Identifying sensitive data

This section contains the following topics:

- [Discovering your sensitive data](#)
- [Out of the box profiling settings](#)
- [ASDD standard profile set](#)
- [Standard profile set expressions](#)
- [Legacy profile set expressions](#)
- [Configuring profile sets](#)
- [Managing domains](#)
- [Managing classifiers](#)
- [Managing expressions](#)
- [ASDD profile set import and export](#)
- [Creating a profiling job](#)
- [Running a profiling job](#)
- [Reporting profiling results](#)
- [ASDD features and support](#)

Discovering your sensitive data

Overview

After connecting data to the masking service, the next step is to discover which of the data should be secured. This process is referred to as *sensitive data discovery*, or *profiling* throughout the product documentation.

Once a rule set has been [created](#), profiling is done by [Managing rule sets](#) and [running](#) a profiling job for that rule set. A profiling job examines the metadata, such as column names and types, and potentially the data itself, to determine which columns or fields contain sensitive information. Upon determining that a data item is sensitive, the profiler assigns the matching domain and associated masking algorithm to the column or field. A profiling job covers only those tables and files present in the rule set; any new objects accessible through the defined connector will not be discovered and must be manually added to the rule set.

The Continuous Compliance product currently ships with two distinct profiling implementations: the new Automated Sensitive Data Discovery (ASDD) profiler and the legacy profiler. The content of the profile set determines which implementation will be chosen when a profiling job is run. The [ASDD profiler supports](#) a wider range of logic for detecting sensitive fields and improved data inspection logic for databases. However, at this time, ASDD profiling is limited to only specific database variants.

Concepts

Profile set

The *Profile Set* chosen defines the logic that will be used to determine which columns or fields in the rule set contain sensitive information. A profile set may contain a set of *search expression* and *type expressions*, or a set of *classifiers*, that define the recognition logic for the legacy or ASDD profiler, respectively. As each expression or classifier is associated with a *domain*, the composition of the profile set determines which types of sensitive data may be detected by a profiling job use a particular profile set. Several [built-in profile sets](#) are available by default.

Domain

A domain represents a particular type of sensitive information, such as first name or tax ID number. Based on the detection logic in the profile set, a profile job may assign a domain to a particular field or column in the rule set; when this occurs, the default masking algorithm defined for that domain will also be assigned. The domain mechanism helps to ensure that the same masking algorithm is applied consistently across rule sets whenever a particular type of sensitive data is discovered.

Level - column or data

The term Level is used for search expressions to indicate whether the data itself is examined, or if profiling is done based only on the field or column name and type. Examining the data is more time-consuming than examining metadata alone, as the profiling job must retrieve data from the data source.

Classifier

A classifier defines a specific piece of logic for recognizing sensitive data. Classifiers may only be used with the ASDD profiler. Classifiers use a framework and instance model, similar to algorithms. A framework represents a particular software module for detecting sensitive information, while an instance provides the configuration for a framework and associates it with a particular domain. The pre-built *ASDD Standard* profile set includes a number of classifier instance definitions. It is possible to create additional instances using the API client.

The following classifier frameworks are available:

- **PATH** - Examines the path to the data in question and applies regular expression and/or exact match logic to match domains. For databases, the path includes the table and column name.
- **TYPE** - Uses the data type and length of a field or column to reject possible domain matches. Supported types are String, Number, Date and Binary.
- **REGEX** - Matches the data itself using regular expressions to match or reject domains.
- **LIST** - Checks whether data values are present in a list of value to match or reject domains.

Of these frameworks, **PATH** and **TYPE** operate at the column level, while **REGEX** and **LIST** operate at the data level. It is not currently possible to install additional classifier frameworks.

Search expression

A search expression defines a regular expression (regex) that will be used to match data to a domain. How the regex is applied depends on the value chosen for level - column-level expressions are matched against the field or column name, while data-level expressions are matched against the data values themselves. Every legacy, built-in profile set includes a number of column-level search expressions

designed to identify common sensitive data types (SSN, Name, Addresses, etc). The pre-built profile sets do not include any data level expressions by default, but some [data level expressions](#) are included (but not part of any profile set) that may be added to user-created profile sets. You also have the ability to create additional search expressions.

Type expression

A type expression defines a constraint limiting matches for a particular domain to a particular set of data types, with an optional minimum length for each type. For example, matches for the FIRST_NAME domain may be limited to only string columns with a length of 8 characters or more. Supported types are STRING, NUMBER, DATE, and BINARY. The [Standard](#) profile set includes type expressions for most domains, and more may be created if desired.

Out of the box profiling settings

The Delphix Platform comes out of the box with recognition logic to help you discover over 30 types of sensitive data (account numbers, addresses, etc.). This logic is organized into a number of pre-built profile sets that can be easily applied to a rule set when a profile job is created.

ASDD standard profile set

This is the recommended profiler set for the ASDD profiler and should be preferred for all <https://delphixdocs.atlassian.net/wiki/spaces/CC/pages/9962933/ASDD+features+and+support> by the ASDD Profiler. This profile set has the widest range of classification logic, including classifiers for all logic in the legacy **Standard** profile set, as well as data-level classifiers for a number of domains. It includes value list classifiers capable of detecting several domains, such as FIRST_NAME and LAST_NAME, even when column names are not meaningful. Data level detection is limited to English language values.

The classifiers present in the ASDD Standard profile set are described in the [ASDD Standard Profile Set](#) section.

Standard profile set

This is the recommended profile set for the legacy profiler. It contains column-level search and type expressions appropriate for detecting a wide range of sensitive information.

The column and type expressions used in this profile set are described in the [Standard Profile Set Expressions](#) section.

Legacy profile sets

The legacy profile sets are provided for backward compatibility, specifically, to provide consistent results for pre-existing profiling jobs. For other uses, the *Standard* profile set described above is preferred. The legacy profile sets do not contain any type expressions to restrict matching based on the column type.

These profile sets are:

- Financial - Legacy
- HIPAA - Legacy

The expressions used by these profile sets are described in the [Legacy Profile Set Expressions](#) section.

ASDD standard profile set

As of the 11.0.0.0 release, administrators control when upgrades of the **ASDD Standard Profile Set** occur. For every Delphix Engine release, the ASDD Standard Profile Set will be made available to customers in the Automated Sensitive Data Discovery directory folder on the [Delphix Download site](#). A README is also present, which includes a changelog and usage instructions. Additionally, as of 11.0.0.0, the contents of the ASDD Standard Profile Set may be customized. To recover or upgrade the ASDD Standard profile set, see the [ASDD Profile Set Import and Export article](#) for more information on the import API endpoint.

For reference, the following JSON listing contains the full configuration for all classifiers present in the ASDD Standard Profile Set as of the 11.0.0.0 release.

```
[
  {
    "domain": "ACCOUNT_NO",
    "name": "Account Number - Path",
    "type": "PATH",
    "properties": {
      "paths": [
        {
          "matchType": "REGEX",
          "matchStrength": 0.67,
          "fieldValue": "(?i)(?>(account|acct|acnt)_?-? ?(number|num|nbr|no|user))$"
        }
      ],
      "rejectStrength": 0
    }
  },
  {
    "domain": "ACCOUNT_NO",
    "name": "Account Number - Type",
    "type": "TYPE",
    "properties": {
      "allowedTypes": [
        {
          "minimumLength": "5",
          "typeName": "String"
        },
        {
          "minimumLength": "5",
          "typeName": "Number"
        }
      ],
      "matchAutoIncrementingColumn": true
    }
  },
  {
    "domain": "ADDRESS",
    "name": "Address Line 1 - Path",
    "type": "PATH",
    "properties": {
      "paths": [
        {
          "matchType": "REGEX",
          "matchStrength": 0.67,
          "fieldValue": "(?i)(?>((street)?_?-? ?addr(ess)?)|street?|addr(ess)?_?-? ?((l(i)?n(e)?)? ?_?(1|))?)?"
        }
      ],
      "rejectStrength": 0
    }
  },
  {
```

```

"domain": "ADDRESS",
"name": "Address Line 1 - Type",
"type": "TYPE",
"properties": {
  "allowedTypes": [
    {
      "minimumLength": "20",
      "typeName": "String"
    }
  ]
},
{
  "domain": "ADDRESS",
  "name": "Address Line 1 - Regex",
  "type": "REGEX",
  "properties": {
    "dataPatterns": [
      {
        "regex": "(?i)(.*[\\s]+b(ou)?l(e)?v(ar)?d[\\s]*.*)|(.*[\\s]+st(reet)?[\\s]*.*)|(.*[\\s]+ave[.]?(nue)?[\\s]*.*)|(.*[\\s]+r(oa)?d[\\s]*.*)|(.*[\\s]+l(a)?n(e)?[\\s]*.*)|(.*[\\s]+cir(cle)?[\\s]*.*)|(.*[\\s]+dr[.]?(ive)?[\\s]*.*)|(.*[\\s]+h(igh)?w(a)?y[\\s]*.*)|(.*[\\s]+r(ou)?t(e)?[\\s]*.*)|(.*[\\s]+c(our)?t[\\s]*.*)|(.*[\\s]+way[\\s]*.*)",
        "matchStrength": 0.7,
        "allowPartialMatch": false
      }
    ],
    "rejectStrength": 0.1
  }
},
{
  "domain": "ADDRESS_LINE2",
  "name": "Address Line 2 - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>((st(reet)?_?-? ?addr(ess)?)|street?|addr(ess)?_?-? ?((l(i)?n(e)? ?_?([2-9]))?))"
      }
    ],
    "rejectStrength": 0
  }
},
{
  "domain": "ADDRESS_LINE2",
  "name": "Address Line 2 - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "20",
        "typeName": "String"
      }
    ]
  }
},
{
  "domain": "ADDRESS_LINE2",
  "name": "Address Line 2 - Regex",
  "type": "REGEX",
  "properties": {
    "dataPatterns": [
      {

```

```

        "regex": "(?i)(.*[\\s]*ap(ar)?t(ment)?[\\s]*.*)|(.*[\\s]*s(ui)?te[\\s]*.*)|(c(are)?[\\s]*[\\\\\\\\]?[\\/]?(o(f)?[\\s]*.*)|(.*[\\s]*b(ui)?ld(in)?g[\\s]*.*)|(.*[\\s]*fl(oor)?[\\s]*.*)|(.*[\\s]*r(oo)?m[\\s]*.*)|(.*[\\s]*unit[\\s]*.*))",
        "matchStrength": 0.7,
        "allowPartialMatch": false
    }
  ],
  "rejectStrength": 0.1
},
{
  "domain": "BANK_ACCOUNT_NO",
  "name": "Bank Account Number - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(bank)?(account|acct|acct)_?-? ?(number|num|nbr|no))$"
      }
    ],
    "rejectStrength": 0
  }
},
{
  "domain": "BANK_ACCOUNT_NO",
  "name": "Bank Account Number - Regex",
  "type": "REGEX",
  "properties": {
    "dataPatterns": [
      {
        "regex": "\\d{5,17}$",
        "matchStrength": 0.05,
        "allowPartialMatch": false
      }
    ],
    "rejectStrength": 0.1
  }
},
{
  "domain": "BANK_ACCOUNT_NO",
  "name": "Bank Account Number - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "5",
        "typeName": "Number"
      }
    ],
    "matchAutoIncrementingColumn": true
  }
},
{
  "domain": "BENEFICIARY_NO",
  "name": "Beneficiary Number - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(bene(ficiary)?_?-? ?(Number|Num|Nbr|No|Id)))$"
      }
    ],
  },

```

```

    "rejectStrength": 0
  }
},
{
  "domain": "BENEFICIARY_NO",
  "name": "Beneficiary Number - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "5",
        "typeName": "Number"
      },
      {
        "minimumLength": "10",
        "typeName": "String"
      }
    ]
  }
},
{
  "domain": "BIOMETRIC",
  "name": "Biometric - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(biometric)$)"
      }
    ],
    "rejectStrength": 0
  }
},
{
  "domain": "BIOMETRIC",
  "name": "Biometric - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "10",
        "typeName": "String"
      },
      {
        "minimumLength": "0",
        "typeName": "Binary"
      }
    ]
  }
},
{
  "domain": "CERTIFICATE_NO",
  "name": "Certificate Number - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>cert(ificate)?_?-? ?id)"
      },
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,

```

```

        "fieldValue": "(?i)(?>(Cert(ificate)?_?-? ?(Number|Num|Nbr|No)))$"
    }
  ],
  "rejectStrength": 0
}
},
{
  "domain": "CERTIFICATE_NO",
  "name": "Certificate Number - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "10",
        "typeName": "String"
      },
      {
        "minimumLength": "5",
        "typeName": "Number"
      }
    ]
  }
}
},
{
  "domain": "CITY",
  "name": "City - List",
  "type": "LIST",
  "properties": {
    "valueLists": [
      {
        "file": "file://us_cities.txt",
        "matchStrength": 1
      }
    ]
  }
}
},
{
  "domain": "CITY",
  "name": "City - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>^(home_?-? ?city|city|city_?-? ?ad?dress?e?)$)"
      },
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>^(address_?-? ?city|city|city_?-? ?address)$)"
      }
    ],
    "rejectStrength": 0
  }
}
},
{
  "domain": "CITY",
  "name": "City - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "10",
        "typeName": "String"
      }
    ]
  }
}

```

```

    ]
  }
},
{
  "domain": "COUNTRY",
  "name": "Country - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)country",
        "allowPartialMatch": false
      }
    ],
    "rejectStrength": 0
  }
},
{
  "domain": "COUNTRY",
  "name": "Country - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "15",
        "typeName": "String"
      }
    ]
  }
},
{
  "domain": "COUNTRY",
  "name": "Country - List",
  "type": "LIST",
  "properties": {
    "valueLists": [
      {
        "file": "file://countries.txt",
        "matchStrength": 1
      }
    ],
    "rejectStrength": 0.5
  }
},
{
  "domain": "COUNTY",
  "name": "County - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(county)$)"
      }
    ],
    "rejectStrength": 0
  }
},
{
  "domain": "COUNTY",
  "name": "County - Type",
  "type": "TYPE",
  "properties": {

```

```

    "allowedTypes": [
      {
        "minimumLength": "10",
        "typeName": "String"
      }
    ]
  },
  {
    "domain": "CREDIT CARD",
    "name": "Credit Card Number - Path",
    "type": "PATH",
    "properties": {
      "paths": [
        {
          "matchType": "REGEX",
          "matchStrength": 0.67,
          "fieldValue": "(?i)(?>credit_?-? ?card_?-? ?(number|num|nbr|no))$"
        },
        {
          "matchType": "REGEX",
          "matchStrength": 0.67,
          "fieldValue": "(?i)(?>card_?-? ?(number|num|nbr|no))$"
        }
      ],
      "rejectStrength": 0
    }
  },
  {
    "domain": "CREDIT CARD",
    "name": "Credit Card Number - Regex",
    "type": "REGEX",
    "properties": {
      "dataPatterns": [
        {
          "regex": "(?:3[47][0-9]{2}[0-9]{6}[0-9]{4})",
          "allowPartialMatch": false,
          "checksumType": "LUHN",
          "dataCleanRegex": "[ -.]",
          "matchStrength": 0.9
        },
        {
          "regex": "(?:4[0-9]{12}(?:[0-9]{3})?(?:[0-9]{3})?)",
          "allowPartialMatch": false,
          "checksumType": "LUHN",
          "dataCleanRegex": "[ -.]",
          "matchStrength": 0.9
        },
        {
          "regex": "(?:5[1-5][0-9]{2}|222[1-9]|22[3-9][0-9]|2[3-6][0-9]{2}|27[01][0-9]|2720)[0-9]{12}",
          "allowPartialMatch": false,
          "checksumType": "LUHN",
          "dataCleanRegex": "[ -.]",
          "matchStrength": 0.9
        },
        {
          "regex": "(?:2131|1800|35[0-9]{3})[0-9]{11}",
          "allowPartialMatch": false,
          "checksumType": "LUHN",
          "dataCleanRegex": "[ -.]",
          "matchStrength": 0.9
        },
        {
          "regex": "3(?:0[0-5,9]|6[0-9])[0-9]{11}|3[89][0-9]{12}?(?:[0-9]{1,3})?",
          "allowPartialMatch": false,

```

```

        "checksumType": "LUHN",
        "dataCleanRegex": "[ -.]",
        "matchStrength": 0.9
    },
    {
        "regex": "6(?:011|5[0-9][0-9])[0-9]{2}|4[4-9][0-9]{3}|2212[6-9]|221[3-9][0-9]|22[2-8]
[0-9]{2}|229[0-1][0-9]|2292[0-5])[0-9]{10}?(?:[0-9]{3})?",
        "allowPartialMatch": false,
        "checksumType": "LUHN",
        "dataCleanRegex": "[ -.]",
        "matchStrength": 0.9
    }
],
"rejectStrength": 0.1
}
},
{
    "domain": "CREDIT_CARD",
    "name": "Credit Card Number - Type",
    "type": "TYPE",
    "properties": {
        "allowedTypes": [
            {
                "minimumLength": "15",
                "typeName": "Number"
            },
            {
                "minimumLength": "15",
                "typeName": "String"
            }
        ]
    }
},
{
    "domain": "CUSTOMER_NO",
    "name": "Customer Number - Path",
    "type": "PATH",
    "properties": {
        "paths": [
            {
                "matchType": "REGEX",
                "matchStrength": 0.67,
                "fieldValue": "(?i)(?>(cust(omer|mr)? ?_?-?(num(ber)?|nbr|no))$"
            }
        ]
    },
    "rejectStrength": 0
}
},
{
    "domain": "CUSTOMER_NO",
    "name": "Customer Number - Type",
    "type": "TYPE",
    "properties": {
        "allowedTypes": [
            {
                "minimumLength": "5",
                "typeName": "String"
            },
            {
                "minimumLength": "5",
                "typeName": "Number"
            }
        ]
    }
},
{

```

```

"domain": "DOB",
"name": "Date of Birth - Path",
"type": "PATH",
"properties": {
  "paths": [
    {
      "matchType": "REGEX",
      "matchStrength": 0.67,
      "fieldValue": "(?i)(?>dob|dtofb|(day|date?|dt)_?-?(of)?_?(birth))$"
    },
    {
      "matchType": "REGEX",
      "matchStrength": 0.67,
      "fieldValue": "(?i)(?>b(irth)?_?-? ?(date|day|dt))$"
    },
    {
      "matchType": "REGEX",
      "matchStrength": 0.67,
      "fieldValue": "(?i)(?>(adm(it|ission)?|tr(ea)?t(ment)?_?-?|ds|disc(h|harge))_? ?(date|day|
dt))$"
    }
  ],
  "rejectStrength": 0
}
},
{
  "domain": "DOB",
  "name": "Date of Birth - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "6",
        "typeName": "String"
      },
      {
        "typeName": "Date"
      }
    ]
  }
}
},
{
  "domain": "DRIVING_LC",
  "name": "Drivers License - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(drivers?|lic(ense)?_?-? ?(number|num|nbr|no))$"
      }
    ],
    "rejectStrength": 0
  }
}
},
{
  "domain": "DRIVING_LC",
  "name": "Drivers License - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "10",
        "typeName": "Number"
      }
    ],
  }
}

```

```

    {
      "minimumLength": "10",
      "typeName": "String"
    }
  ]
}
},
{
  "domain": "DRIVING_LC",
  "name": "Drivers License - Regex",
  "type": "REGEX",
  "properties": {
    "dataPatterns": [
      {
        "note": "One alpha followed by digits (short): AZ 1+8, CA 1+7, HI 1+8, IN 1+9, KS 1+8, MA 1+9, MO 1+5..9, MT 1+9, NE 1+6..8, NV 1+8, OH 1+4..8, OK 1+9, OR 1+6..7, RI 1+6, UT 4-10, VA 1+8..11, VT 1+8, WV 1+6",
        "regex": "[A-Z][0-9]{4,9}",
        "matchStrength": 0.15,
        "allowPartialMatch": false
      },
      {
        "note": "One alpha followed by digits (long): FL 1+12, MD 1+12, MI 1+12, MN 1+12, NJ 1+14, UT 4-10, VA 1+8..11",
        "regex": "[A-Z][0-9]{10,14}",
        "matchStrength": 0.3,
        "allowPartialMatch": false
      },
      {
        "note": "One alpha followed by digits with dashes (short): KS 1+8, KY 1+8, VA 1+8..11",
        "dataCleanRegex": "[^-]",
        "regex": "[A-Z][0-9]{8,9}",
        "matchStrength": 0.15,
        "allowPartialMatch": false
      },
      {
        "note": "One alpha followed by digits with dashes (long): FL 1+12, IL 1+11 or 1+12, MD 1+12, MN 1+12, VA 1+8..11, WI 1+13",
        "dataCleanRegex": "[^-]",
        "regex": "[A-Z][0-9]{10,13}",
        "matchStrength": 0.3,
        "allowPartialMatch": false
      },
      {
        "note": "Two alpha followed by digits: OH 2+3..7, WV 2+7",
        "regex": "[A-Z]{2}[0-9]{3,7}",
        "matchStrength": 0.2,
        "allowPartialMatch": false
      },
      {
        "note": "Digits followed by alpha: ME 7+1, MO 8+2 or 9+1, VT 7+1",
        "regex": "[0-9]{7}[A-Z]|[0-9]{8}[A-Z]{2}|[0-9]{9}[A-Z]",
        "matchStrength": 0.2,
        "allowPartialMatch": false
      },
      {
        "note": "Digits and spaces: NM 3-3-3, NY 3-3-3, PA 2-3-3",
        "regex": "[0-9]{2,3} [0-9]{3} [0-9]{3}",
        "matchStrength": 0.1,
        "allowPartialMatch": false
      },
      {
        "note": "Digit-alpha-digit: IA 3+2+4, NH 2+3+5",
        "regex": "[0-9]{3}[A-Z]{2}[0-9]{4}|[0-9]{2}[A-Z]{3}[0-9]{5}",
        "matchStrength": 0.3,
        "allowPartialMatch": false
      }
    ]
  }
}

```

```

    },
    {
      "note": "Digits only: AL 7-8, AK 7, AZ 9, AR 9, CT 9, DE 7, DC 7, GA 9, HI 9, IA 9, LA 8,
ME 7-8, MA 9, MO 9, MT 9 or 13-14, NV 9-10 or 12, NY 9, NM 8-9, NC 1-12, ND 9, OH 8, OK 9, RI 7, SC
5-11, SD 6-10 or 12, TN 7-9, TX 7-8, VA 9, WV 7, WY 9",
      "regex": "([0-9]{6,14})",
      "matchStrength": 0,
      "allowPartialMatch": false
    },
    {
      "note": "CO",
      "regex": "[0-9]{2}-[0-9]{3}-[0-9]{4}",
      "matchStrength": 0.1,
      "allowPartialMatch": false
    },
    {
      "note": "ID",
      "regex": "[A-Z]{2}[0-9]{6}[A-Z]",
      "matchStrength": 0.3,
      "allowPartialMatch": false
    },
    {
      "note": "IN",
      "regex": "[0-9]{2}-[0-9]{4}-[0-9]{4}",
      "matchStrength": 0.1,
      "allowPartialMatch": false
    },
    {
      "note": "MA - SA + 7 digits",
      "regex": "SA[0-9]{7}",
      "matchStrength": 0.5,
      "allowPartialMatch": false
    },
    {
      "note": "MI - 1 alpha + 12 digits with spaces",
      "regex": "[A-Z]([0-9]{3}){4}",
      "matchStrength": 0.5,
      "allowPartialMatch": false
    },
    {
      "note": "MS - looks just like SSN with dashes",
      "regex": "([0-9]{3}-[0-9]{2}-[0-9]{4})",
      "matchStrength": 0,
      "allowPartialMatch": false
    },
    {
      "note": "MO - not covered elsewhere",
      "regex": "[0-9]{3}[A-Z][0-9]{6}|[A-Z][0-9]{6}R",
      "matchStrength": 0.3,
      "allowPartialMatch": false
    },
    {
      "note": "NH - NHL + 8 digits",
      "regex": "NHL[0-9]{8}",
      "matchStrength": 0.5,
      "allowPartialMatch": false
    },
    {
      "note": "ND - 3 letters, 6 digits with spaces",
      "regex": "[A-Z]{3} ?[0-9]{2} ?[0-9]{4}",
      "matchStrength": 0.3,
      "allowPartialMatch": false
    },
    {
      "note": "NJ 1 alpha + 14 digits with spaces",
      "dataCleanRegex": "[-]",

```

```

    "regex": "[A-Z][0-9]{4} [0-9]{5} [0-9]{5}",
    "matchStrength": 0.4,
    "allowPartialMatch": false
  },
  {
    "note": "WA - old - very broad will match any 12 position alpha string",
    "regex": "[A-Z]([A-Z]{4}|[A-Z]{3}[*]|[A-Z]{2}[*]{2}|[A-Z]{1}[*]{3})[A-Z]{2}[0-9A-Z]{5}",
    "matchStrength": 0,
    "allowPartialMatch": false
  },
  {
    "note": "WA - new",
    "regex": "WDL[0-9A-Z]{9}",
    "matchStrength": 0.3,
    "allowPartialMatch": false
  },
  {
    "note": "WY - 9 digits with a dash",
    "regex": "[0-9]{6}-[0-9]{3}",
    "matchStrength": 0.1,
    "allowPartialMatch": false
  }
],
"rejectStrength": 0.75
}
},
{
  "domain": "EMAIL",
  "name": "Email Address - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(cust|customer|partner|home|private|def|default)_?-? ?(email)_?-? ?
(address|)"
      },
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(email_?-? ?)(addr?e?s?s?)?)$"
      }
    ],
    "rejectStrength": 0
  }
},
{
  "domain": "EMAIL",
  "name": "Email Address - Regex",
  "type": "REGEX",
  "properties": {
    "dataPatterns": [
      {
        "regex": "[A-Z0-9.!#$%&'*/+=?^_[]~}{1,64}@(?=. {1,255}$)[A-Z0-9-]+(?:\\. [A-Z0-9-]+)*",
        "matchStrength": 0.9,
        "allowPartialMatch": false
      }
    ],
    "rejectStrength": 0.1
  }
},
{
  "domain": "EMAIL",
  "name": "Email Address - Type",
  "type": "TYPE",

```

```

"properties": {
  "allowedTypes": [
    {
      "minimumLength": "20",
      "typeName": "String"
    }
  ]
}
},
{
  "domain": "FIRST_NAME",
  "name": "First Name - List",
  "type": "LIST",
  "properties": {
    "valueLists": [
      {
        "file": "file://us_first.txt",
        "matchStrength": 1
      },
      {
        "file": "file://de_first.txt",
        "matchStrength": 1
      },
      {
        "file": "file://ch_first.txt",
        "matchStrength": 1
      }
    ]
  }
},
{
  "domain": "FIRST_NAME",
  "name": "First Name - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(mid(dle)?_?-? ?(na?me?))(_?-?user)?)$"
      },
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(f(first)?_?-? ?(na?me?))(_?-?user)?)$"
      }
    ],
    "rejectStrength": 0
  }
},
{
  "domain": "FIRST_NAME",
  "name": "First Name - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "10",
        "typeName": "String"
      }
    ]
  }
},
{
  "domain": "FULL_NAME",
  "name": "Full Name - Path",

```

```

"type": "PATH",
"properties": {
  "paths": [
    {
      "matchType": "REGEX",
      "matchStrength": 0.67,
      "fieldValue": "(?i)?>((fu?l?l|whole|user)([-_ ]*)?(na?me?))$"
    }
  ],
  "rejectStrength": 0
}
},
{
  "domain": "FULL_NAME",
  "name": "Full Name - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "20",
        "typeName": "String"
      }
    ]
  }
},
{
  "domain": "IBAN",
  "name": "IBAN - Regex",
  "type": "REGEX",
  "properties": {
    "dataPatterns": [
      {
        "note": "Fetched from: 'https://www.swift.com/standards/data-standards/iban-international-bank-account-number', List of countries in order for regex formats listed below: Andorra, United Arab Emirates (The), Albania, Austria, Azerbaijan, Bosnia and Herzegovina, Belgium, Bulgaria, Bahrain, Burundi, Brazil, Republic of Belarus, Switzerland, Costa Rica, Cyprus, Czechia, Germany, Djibouti, Denmark, Dominican Republic, Estonia, Egypt, Spain, Finland, Faroe Islands, France, United Kingdom, Georgia, Gibraltar, Greenland, Greece, Guatemala, Croatia, Ireland, Israel, Iraq, Iceland, Italy, Jordan, Kuwait, Kazakhstan, Lebanon, Saint Lucia, Liechtenstein, Lithuania, Luxembourg, Latvia, Libya, Monaco, Moldova, Montenegro, Macedonia, Mauritania, Malta, Mauritius, Netherlands, Norway, Pakistan, Poland, Palestine, State of, Portugal, Qatar, Romania, Serbia, Russia, Saudi Arabia, Seychelles, Sudan, Sweden, Slovenia, Slovakia, San Marino, Somalia, Sao Tome and Principe, El Salvador, Timor-Leste, Tunisia, Turkey, Ukraine, Vatican City State, Virgin Islands, Kosovo",
        "regex": "(AD[0-9]{10}[A-Z0-9]{12})|(AE[0-9]{21})|(AL[0-9]{10}[A-Z0-9]{16})|(AT[0-9]{18})|(AZ[0-9]{2}[A-Z]{4}[A-Z0-9]{20})|(BA[0-9]{18})|(BE[0-9]{14})|(BG[0-9]{2}[A-Z]{4}[0-9]{6}[A-Z0-9]{8})|(BH[0-9]{2}[A-Z]{4}[A-Z0-9]{14})|(BI[0-9]{25})|(BR[0-9]{25}[A-Z]{1}[A-Z0-9]{1})|(BY[0-9]{2}[A-Z0-9]{4}[0-9]{4}[A-Z0-9]{16})|(CH[0-9]{7}[A-Z0-9]{12})|(CR[0-9]{20})|(CY[0-9]{10}[A-Z0-9]{16})|(CZ[0-9]{22})|(DE[0-9]{20})|(DJ[0-9]{25})|(DK[0-9]{16})|(DO[0-9]{2}[A-Z0-9]{4}[0-9]{20})|(EE[0-9]{18})|(EG[0-9]{27})|(ES[0-9]{22})|(FI[0-9]{16})|(FO[0-9]{16})|(FR[0-9]{12}[A-Z0-9]{11}[0-9]{2})|(GB[0-9]{2}[A-Z]{4}[0-9]{14})|(GE[0-9]{2}[A-Z]{2}[0-9]{16})|(GI[0-9]{2}[A-Z]{4}[A-Z0-9]{15})|(GL[0-9]{16})|(GR[0-9]{9}[A-Z0-9]{16})|(GT[0-9]{2}[A-Z0-9]{24})|(HR[0-9]{19})|(HU[0-9]{26})|(IE[0-9]{2}[A-Z]{4}[0-9]{14})|(IL[0-9]{21})|(IQ[0-9]{2}[A-Z]{4}[0-9]{15})|(IS[0-9]{24})|(IT[0-9]{2}[A-Z]{1}[0-9]{10}[A-Z0-9]{12})|(JO[0-9]{2}[A-Z]{4}[0-9]{4}[A-Z0-9]{18})|(KW[0-9]{2}[A-Z]{4}[A-Z0-9]{22})|(KZ[0-9]{5}[A-Z0-9]{13})|(LB[0-9]{6}[A-Z0-9]{20})|(LC[0-9]{2}[A-Z]{4}[A-Z0-9]{24})|(LI[0-9]{7}[A-Z0-9]{12})|(LT[0-9]{18})|(LU[0-9]{5}[A-Z0-9]{13})|(LV[0-9]{2}[A-Z]{4}[A-Z0-9]{13})|(LY[0-9]{23})|(MC[0-9]{12}[A-Z0-9]{11}[0-9]{2})|(MD[0-9]{2}[A-Z0-9]{20})|(ME[0-9]{20})|(MK[0-9]{5}[A-Z0-9]{10}[0-9]{2})|(MR[0-9]{25})|(MT[0-9]{2}[A-Z]{4}[0-9]{5}[A-Z0-9]{18})|(MU[0-9]{2}[A-Z]{4}[0-9]{19}[A-Z]{3})|(NL[0-9]{2}[A-Z]{4}[0-9]{10})|(NO[0-9]{13})|(PK[0-9]{2}[A-Z]{4}[A-Z0-9]{16})|(PL[0-9]{26})|(PS[0-9]{2}[A-Z]{4}[A-Z0-9]{21})|(PT[0-9]{23})|(QA[0-9]{2}[A-Z]{4}[A-Z0-9]{21})|(RO[0-9]{2}[A-Z]{4}[A-Z0-9]{16})|(RS[0-9]{20})|(RU[0-9]{31})|(SA[0-9]{4}[A-Z0-9]{18})|(SC[0-9]{2}[A-Z]{4}[0-9]{20}[A-Z]{3})|(SD[0-9]{16})|(SE[0-9]{22})|(SI[0-9]{17})|(SK[0-9]{22})|(SM[0-9]{2}[A-Z]{1}[0-9]{10}[A-Z0-9]{12})|(SO[0-9]{21})|(ST[0-9]{23})|(SV[0-9]{2}[A-Z]{4}[0-9]{20})|(TL[0-9]{21})|(TN[0-9]{22})|(TR[0-9]{8}[A-Z0-9]{16})|(UA[0-9]{8}[A-Z0-9]{19})|(VA[0-9]{20})|(VG[0-9]{2}[A-Z]{4}[0-9]{16})|(XK[0-9]{18})",
        "allowPartialMatch": false,

```

```

        "checksumType": "MOD97",
        "matchStrength": 0.9
    }
  ],
  "rejectStrength": 0.1
}
},
{
  "domain": "IP ADDRESS",
  "name": "IP Address - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(ip_?-? ?adre?s?s?))$"
      }
    ],
    "rejectStrength": 0
  }
},
{
  "domain": "IP ADDRESS",
  "name": "IP Address - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "10",
        "typeName": "String"
      }
    ]
  }
},
{
  "domain": "LAST_NAME",
  "name": "Last Name - List",
  "type": "LIST",
  "properties": {
    "valueLists": [
      {
        "file": "file://us_last.txt",
        "matchStrength": 1
      },
      {
        "file": "file://de_last.txt",
        "matchStrength": 1
      },
      {
        "file": "file://ch_last.txt",
        "matchStrength": 1
      }
    ]
  }
},
{
  "domain": "LAST_NAME",
  "name": "Last Name - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>((l(as)?t)_?-? ?(na?me?))(_?-?user)?)"
      }
    ]
  }
}

```

```

    },
    {
      "matchType": "REGEX",
      "matchStrength": 0.67,
      "fieldValue": "(?i)(?>(sur) ?_?-? ?(name)?_?-? ?(no|id|str|value|))"
    }
  ],
  "rejectStrength": 0
}
},
{
  "domain": "LAST_NAME",
  "name": "Last Name - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "10",
        "typeName": "String"
      }
    ]
  }
},
{
  "domain": "PO_BOX",
  "name": "PO Box - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(p.?o.?_?-? ?box|post_?-? ?office_?-? ?box ?_?-?)(number|num|nbr|
no)?$)"
      }
    ]
  },
  "rejectStrength": 0
}
},
{
  "domain": "PO_BOX",
  "name": "PO Box - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "4",
        "typeName": "String"
      },
      {
        "minimumLength": "4",
        "typeName": "Number"
      }
    ]
  }
},
{
  "domain": "PASSWORD",
  "name": "Password - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(pass) ?_?-??(word)?_?-? ?(word|nbr|no|id|value|))"
      }
    ]
  }
}

```

```

    }
  ],
  "rejectStrength": 0
}
},
{
  "domain": "PASSWORD",
  "name": "Password - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "6",
        "typeName": "String"
      }
    ]
  }
},
{
  "domain": "ZIP",
  "name": "Postcode - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(zip|post|postal)_?-? ?(co?de?)|(zip))"
      }
    ]
  },
  "rejectStrength": 0
}
},
{
  "domain": "ZIP",
  "name": "Postcode - Regex",
  "type": "REGEX",
  "properties": {
    "dataPatterns": [
      {
        "regex": "[0-9]{5}(?:-[0-9]{4})$",
        "matchStrength": 0.7
      },
      {
        "regex": "[0-9]{5}$",
        "matchStrength": 0.2
      }
    ]
  },
  "rejectStrength": 0.1
}
},
{
  "domain": "ZIP",
  "name": "Postcode - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "4",
        "typeName": "Number"
      },
      {
        "minimumLength": "4",
        "typeName": "String"
      }
    ]
  }
}
]

```

```

    }
  },
  {
    "domain": "PRECINCT",
    "name": "Precinct - Path",
    "type": "PATH",
    "properties": {
      "paths": [
        {
          "matchType": "REGEX",
          "matchStrength": 0.67,
          "fieldValue": "(?i)(?>precinct|prcnct)$"
        }
      ],
      "rejectStrength": 0
    }
  },
  {
    "domain": "PRECINCT",
    "name": "Precinct - Type",
    "type": "TYPE",
    "properties": {
      "allowedTypes": [
        {
          "minimumLength": "0",
          "typeName": "String"
        },
        {
          "minimumLength": "0",
          "typeName": "Number"
        }
      ]
    }
  },
  {
    "domain": "RECORD_NO",
    "name": "Record Number - Path",
    "type": "PATH",
    "properties": {
      "paths": [
        {
          "matchType": "REGEX",
          "matchStrength": 0.67,
          "fieldValue": "(?i)(?>(rec|record)_?(number|num|nbr|no))$"
        }
      ],
      "rejectStrength": 0
    }
  },
  {
    "domain": "RECORD_NO",
    "name": "Record Number - Type",
    "type": "TYPE",
    "properties": {
      "allowedTypes": [
        {
          "minimumLength": "5",
          "typeName": "String"
        },
        {
          "minimumLength": "5",
          "typeName": "Number"
        }
      ]
    }
  },
  },

```

```

{
  "domain": "SCHOOL_NM",
  "name": "School Name - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>school_?-?na?me?)$"
      }
    ],
    "rejectStrength": 0
  }
},
{
  "domain": "SCHOOL_NM",
  "name": "School Name - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "20",
        "typeName": "String"
      }
    ]
  }
},
{
  "domain": "SECURITY_CODE",
  "name": "Security Code - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>se?cu?r(i?ty)?_?co?de?)$"
      }
    ],
    "rejectStrength": 0
  }
},
{
  "domain": "SECURITY_CODE",
  "name": "Security Code - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "5",
        "typeName": "String"
      },
      {
        "minimumLength": "5",
        "typeName": "Number"
      },
      {
        "minimumLength": "0",
        "typeName": "Binary"
      }
    ]
  }
},
{
  "domain": "SERIAL_NO",

```

```

"name": "Serial Number - Path",
"type": "PATH",
"properties": {
  "paths": [
    {
      "matchType": "REGEX",
      "matchStrength": 0.67,
      "fieldValue": "(?i)(?>(ser(ial)?)-?-? ?(number|num|nbr|no))$"
    }
  ],
  "rejectStrength": 0
}
},
{
  "domain": "SERIAL_NO",
  "name": "Serial Number - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "0",
        "typeName": "Number"
      },
      {
        "minimumLength": "0",
        "typeName": "String"
      }
    ]
  }
}
},
{
  "domain": "SIGNATURE",
  "name": "Signature - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(signature)$)"
      }
    ],
    "rejectStrength": 0
  }
}
},
{
  "domain": "SIGNATURE",
  "name": "Signature - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "0",
        "typeName": "String"
      },
      {
        "minimumLength": "0",
        "typeName": "Binary"
      }
    ]
  }
}
},
{
  "domain": "SSN",
  "name": "Social Security Number - Path",
  "type": "PATH",

```

```

"properties": {
  "paths": [
    {
      "matchType": "REGEX",
      "matchStrength": 0.67,
      "fieldValue": "(?i)(?>(ssn$|social_?-? ?security_?-? ?(number|num|nbr|no|code|id))$)"
    }
  ],
  "rejectStrength": 0
}
},
{
  "domain": "SSN",
  "name": "Social Security Number - Regex",
  "type": "REGEX",
  "properties": {
    "dataPatterns": [
      {
        "regex": "(?!000)(?!666)[0-8]\\d{2}([- ])(?!00)\\d{2}\\1(?!0000)\\d{4}",
        "allowPartialMatch": false,
        "matchStrength": 0.7
      },
      {
        "regex": "[9]\\d{2}([- ])(?!6[6-9])(?!89)([5-9][0-9])\\1\\d{4}",
        "allowPartialMatch": false,
        "matchStrength": 0.7
      },
      {
        "regex": "(?!000)(?!666)[0-8]\\d{2}(?!00)\\d{2}(?!0000)\\d{4}",
        "allowPartialMatch": false,
        "matchStrength": 0.1
      },
      {
        "regex": "[9]\\d{2}(?!6[6-9])(?!89)([5-9][0-9])\\d{4}",
        "allowPartialMatch": false,
        "matchStrength": 0.1
      }
    ],
    "rejectStrength": 0.1
  }
},
{
  "domain": "SSN",
  "name": "Social Security Number - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "9",
        "typeName": "Number"
      },
      {
        "minimumLength": "9",
        "typeName": "String"
      }
    ]
  }
},
{
  "domain": "TAX_ID",
  "name": "Tax ID - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",

```

```

        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(tax)_?-? ?(id(ent)?_?-? ?((co?de?)|(number|num|nbr|no)))?)$"
    },
    {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>tin$)"
    }
],
"rejectStrength": 0
}
},
{
    "domain": "TAX_ID",
    "name": "Tax ID - Type",
    "type": "TYPE",
    "properties": {
        "allowedTypes": [
            {
                "minimumLength": "6",
                "typeName": "String"
            },
            {
                "minimumLength": "6",
                "typeName": "Number"
            }
        ]
    }
},
{
    "domain": "TELEPHONE_NO",
    "name": "Telephone Number - Path",
    "type": "PATH",
    "properties": {
        "paths": [
            {
                "matchType": "REGEX",
                "matchStrength": 0.67,
                "fieldValue": "(?i)(?>(phone|contact|tel|fax)_?-? ?)(number|num|nbr|no)?$"
            }
        ]
    },
    "rejectStrength": 0
}
},
{
    "domain": "TELEPHONE_NO",
    "name": "Telephone Number - Regex",
    "type": "REGEX",
    "properties": {
        "dataPatterns": [
            {
                "regex": "((\\+?1|001)[- .]?)?\\((\\[0-9]{3}\\)\\)[- .]?|[0-9]{3}[- . ]?[0-9]{3}[- . ]?[0-9]{4}",
                "allowPartialMatch": false,
                "matchStrength": 0.7
            },
            {
                "regex": "\\+(?:[0-9][ .()\\/-]?)?{6,14}[0-9]",
                "allowPartialMatch": false,
                "matchStrength": 0.5
            },
            {
                "regex": "[0-9]{5,17}",
                "allowPartialMatch": false,
                "dataCleanRegex": "[+ -.()\\/-]",
                "matchStrength": 0.05
            }
        ]
    }
}

```

```

    ],
    "rejectStrength": 0.1
  }
},
{
  "domain": "TELEPHONE_NO",
  "name": "Telephone Number - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "7",
        "typeName": "Number"
      },
      {
        "minimumLength": "7",
        "typeName": "String"
      }
    ]
  }
},
{
  "domain": "US_PASSPORT",
  "name": "US Passport - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(passport)[_ -]?(number|num|nbr|no)?"
      }
    ],
    "rejectStrength": 0
  }
},
{
  "domain": "US_PASSPORT",
  "name": "US Passport - Regex",
  "type": "REGEX",
  "properties": {
    "dataPatterns": [
      {
        "regex": "[A-Z]\\d{8}",
        "allowPartialMatch": false,
        "matchStrength": 0.2
      },
      {
        "regex": "(?!0{9})\\d{9}",
        "allowPartialMatch": false,
        "matchStrength": 0.1
      }
    ],
    "rejectStrength": 0.1
  }
},
{
  "domain": "US_PASSPORT",
  "name": "US Passport - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "9",
        "typeName": "String"
      }
    ],
  },

```

```

        {
            "minimumLength": "9",
            "typeName": "Number"
        }
    ]
}
},
{
    "domain": "US_STATE",
    "name": "US State - Path",
    "type": "PATH",
    "properties": {
        "paths": [
            {
                "matchType": "REGEX",
                "matchStrength": 0.67,
                "fieldValue": "(?i)state",
                "allowPartialMatch": false
            }
        ],
        "rejectStrength": 0
    }
},
{
    "domain": "US_STATE",
    "name": "US State - Type",
    "type": "TYPE",
    "properties": {
        "allowedTypes": [
            {
                "minimumLength": "14",
                "typeName": "String"
            }
        ]
    }
},
{
    "domain": "US_STATE",
    "name": "US State - List",
    "type": "LIST",
    "properties": {
        "valueLists": [
            {
                "file": "file://us_states_full.txt",
                "matchStrength": 1
            }
        ],
        "rejectStrength": 0.1
    }
},
{
    "domain": "USPS_STATE_CODE",
    "name": "USPS State Code - Path",
    "type": "PATH",
    "properties": {
        "paths": [
            {
                "matchType": "REGEX",
                "matchStrength": 0.67,
                "fieldValue": "(?i)state[ _-]? (cd|code|abbrev)?",
                "allowPartialMatch": false
            }
        ],
        "rejectStrength": 0
    }
},

```

```

{
  "domain": "USPS_STATE_CODE",
  "name": "USPS State Code - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "2",
        "typeName": "String"
      }
    ]
  }
},
{
  "domain": "USPS_STATE_CODE",
  "name": "USPS State Code - List",
  "type": "LIST",
  "properties": {
    "valueLists": [
      {
        "file": "file://us_states.txt",
        "matchStrength": 1
      }
    ],
    "rejectStrength": 0.5
  }
},
{
  "domain": "VIN_NO",
  "name": "Vehicle Identification Number - Path",
  "type": "PATH",
  "properties": {
    "paths": [
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(^vin$|Vehicle_?-? ?Id(entification)?_?-? ?(Number|Num|Nbr|No))$)"
      },
      {
        "matchType": "REGEX",
        "matchStrength": 0.67,
        "fieldValue": "(?i)(?>(vehicle)$)"
      }
    ],
    "rejectStrength": 0
  }
},
{
  "domain": "VIN_NO",
  "name": "Vehicle Identification Number - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "10",
        "typeName": "String"
      }
    ]
  }
},
{
  "domain": "WEB",
  "name": "Web URL - Path",
  "type": "PATH",
  "properties": {
    "paths": [

```

```
    {
      "matchType": "REGEX",
      "matchStrength": 0.67,
      "fieldValue": "(?i)(?>(^url_?-? ?|web_? ?)(addr?s?s?)?)$"
    }
  ],
  "rejectStrength": 0
}
},
{
  "domain": "WEB",
  "name": "Web URL - Type",
  "type": "TYPE",
  "properties": {
    "allowedTypes": [
      {
        "minimumLength": "10",
        "typeName": "String"
      }
    ]
  }
}
}
]Delphix Download site
```

Standard profile set expressions

This section lists all the column and type profile expressions used by the standard profile set included with the product.

Column level expressions

Expression Name	Domain	Expression
Account_Number_V2	ACCOUNT_NO	(?>(account acct acct)?_? ?(number num nbr no user))\$
Address_Line1_V2	ADDRESS	(?>((street_?_? ?address) street address)?_? ?((line)? ?_(1))?)\$
Address_Line2_V2	ADDRESS_LINE2	(?>((street_?_? ?address) street address)?_? ?((line)? ?_(2 3 4 5))?)\$
Beneficiary_NO_V2	BENEFICIARY_NO	(?>(bene(ficiary)?_?_? ?(Number Num Nbr No Id)))\$
Biometric_V2	BIOMETRIC	(?>(biometric))\$
Certificate_Number_V2	CERTIFICATE_NO	(?>(Certificate)?_?_? ?(Number Num Nbr No)))\$
Certificate_ID_V2	CERTIFICATE_NO	(?>certificate)?_?_? ?id)
City_V2	CITY	(?>^(address_?_? ?city city city_?_? ?address)\$)
City_V2_2	CITY	(?>^(address home_?_? ?city city city_?_? ?address?e?)\$)
County_V2	COUNTY	(?>(county))\$

Expression Name	Domain	Expression
Card_Number_V2	CREDIT CARD	(?>card_?-? ?(number num nbr no))\$
Credit_Card_Number_V2	CREDIT CARD	(?>credit_?-? ?card_?-? ?(number num nbr no))\$
Customer_Number_V2	CUSTOMER_NO	(?>(cust(omer mr)?) ?_?-?(num(ber)? nbr no))\$
Birth_Date_V2	DOB	(?>b(irth)?_?-? ?(date day dt))\$
DOB_Date_V2	DOB	(?>dob dtofb (day date? dt)_?-?(of)?_?(birth))\$
Admission_Date_V2	DOB	(?>(adm(it ission)? tr(ea)?t(ment)?_?-? ds disc(h harge))_? ?(date day dt))\$
Drivers_License_Number_V2	DRIVING_LC	(?>(drivers? lic(ense)?)_?-? ?(number num nbr no))\$
Email_V2	EMAIL	(?>(email_?-? ?)(addr?e?s?s?)?)\$
Email_V2_2	EMAIL	(cust customer partner home private def default)_?-? ?(email)_?-? ?(address)
First_Name_V2	FIRST_NAME	(?>(f(irst)?_?-? ?(na?me?))(_?-?user)?)\$
Middle_Name_V2	FIRST_NAME	(?>(mid(dle)?_?-? ?(na?me?))(_?-?user)?)\$
Full_Name_V2	FULL_NAME	(?>((fu?l?l_?-? ? whole_?-? ?)?_?-?(na?me?))(_?-?user)?)\$
IP_Address_V2	IP ADDRESS	(?>(ip_?-? ?addr?e?s?s?)?)\$

Expression Name	Domain	Expression
Last_Name_V2	LAST_NAME	(?>((l(as)?t)?_?-? ?(na?me?))(_?-?user?))\$
Surname_V2	LAST_NAME	(?>(sur) ?_?-? ?(name)?_?-? ?(no id str value))
Password_V2	PASSWORD	(?>(pass) ?_?-??(word)?_?-? ?(word nbr no id value))
PO_Box_V2	PO_BOX	(?>(p.?o.?_?-? ?box post_?-? ?office_?-? ?box ?_?-?) (number num nbr no)?\$)
Precinct_V2	PRECINCT	(?>precinct prcnct)\$
Record_Number_V2	RECORD_NUMBER	(?>(rec record)_?(number num nbr no))\$
School_Name_V2	SCHOOL_NAME	(?>school_?-?na?me?)\$
Security_Code_V2	SECURITY_CODE	(?>se?cu?r(i?ty)?_?co?de?)\$
Serial_Number_V2	SERIAL_NO	(?>(ser(ial)?)_?-? ?(number num nbr no))\$
Signature_V2	SIGNATURE	(?>(signature)\$)
Social_Security_Number_V2	SSN	(?>(ssn\$ social_?-? ?security_?-? ?(number num nbr no code id))\$)
TaxID_Code_or_Number_V2	TAX_ID	(?>(tax)_?-? ?(id(ent)?)_?-? ?((co?de?) (number num nbr no))?)\$
TaxID_Number_V2	TAX_ID	(?>tin\$)

Expression Name	Domain	Expression
Telephone_or_Contact_Number_V2	TELEPHONE_NO	(?>(phone contact tel fax)_?-? ?)(number num nbr no)?\$
Vehicle_V2	VIN_NO	(?>(vehicle)\$)
VIN_NO_V2	VIN_NO	(?>(^vin\$ Vehicle_?-? ?Id(entification)?_?-? ?(Number Num Nbr No))\$)
Web_URL_Address_V2	WEB	(?>(^url_?-? ? web_? ?)(addr?e?s?s?)?)\$
Zip_or_Postal_Code_V2	ZIP	(?>(zip post postal)_?-? ?(co?de?) (zip))

Type expressions

The column data type, if recognized, must match one of the specified types for the domain to be assigned.

Expression Name	Domain	Expression	Minimum Length
ACCOUNT_NO_type_V2	ACCOUNT_NO	Number	5
ACCOUNT_NO_type_V2_2	ACCOUNT_NO	String	5
ADDRESS_type_V2	ADDRESS	String	20
ADDRESS_LINE2_type_V2	ADDRESS_LINE2	String	20
BENEFICIARY_NO_type_V2	BENEFICIARY_NO	String	10
BENEFICIARY_NO_type_V2_2	BENEFICIARY_NO	Number	5
BIOMETRIC_type_V2	BIOMETRIC	String	10
BIOMETRIC_type_V2_2	BIOMETRIC	Binary	None
CERTIFICATE_NO_type_V2	CERTIFICATE_NO	String	10
CERTIFICATE_NO_type_V2_2	CERTIFICATE_NO	Number	5

Expression Name	Domain	Expression	Minimum Length
CITY_type_V2	CITY	String	10
COUNTY_type_V2	COUNTY	String	10
CREDIT_CARD_type_V2	CREDIT_CARD	String	15
CREDIT_CARD_type_V2_2	CREDIT_CARD	Number	15
CUSTOMER_NO_type_V2	CUSTOMER_NO	String	5
CUSTOMER_NO_type_V2_2	CUSTOMER_NO	Number	5
DOB_type_V2	DOB	String	6
DOB_type_V2_2	DOB	Date	None
DRIVING_LC_type_V2	DRIVING_LC	String	10
DRIVING_LC_type_V2_2	DRIVING_LC	Number	10
EMAIL_type_V2	EMAIL	String	20
FIRST_NAME_type_V2	FIRST_NAME	String	10
FULL_NAME_type_V2	FULL_NAME	String	20
IP_ADDRESS_type_V2	IP_ADDRESS	String	10
LAST_NAME_type_V2	LAST_NAME	String	10
PASSWORD_type_V2	PASSWORD	String	6
PO_BOX_type_V2	PO_BOX	String	4
PO_BOX_type_V2_2	PO_BOX	Number	4
PRECINCT_type_V2	PRECINCT	String	None

Expression Name	Domain	Expression	Minimum Length
PRECINCT_type_V2_2	PRECINCT	Number	None
RECORD_NO_type_V2	RECORD_NO	String	5
RECORD_NO_type_V2_2	RECORD_NO	Number	5
SCHOOL_NM_type_V2	SCHOOL_NM	String	20
SECURITY_CODE_type_V2	SECURITY_CODE	String	5
SECURITY_CODE_type_V2_2	SECURITY_CODE	Number	5
SERIAL_NO_type_V2	SERIAL_NO	Number	None
SERIAL_NO_type_V2_2	SERIAL_NO	String	None
SIGNATURE_type_V2	SIGNATURE	String	None
SIGNATURE_type_V2_2	SIGNATURE	Binary	None
SSN_type_V2	SSN	String	None
SSN_type_V2_2	SSN	Number	None
TAX_ID_type_V2	TAX_ID	String	6
TAX_ID_type_V2_2	TAX_ID	Number	6
TELEPHONE_NO_type_V2	TELEPHONE_NO	String	7
TELEPHONE_NO_type_V2_2	TELEPHONE_NO	Number	7
VIN_NO_type_V2	VIN_NO	String	10
WEB_type_V2	WEB	String	10
ZIP_type_V2	ZIP	Number	4

Expression Name	Domain	Expression	Minimum Length
ZIP_type_V2_2	ZIP	String	4

Legacy profile set expressions

This section describes all the column level profile expressions used by the two legacy Profile Sets included with the product, as well as several data level expressions included with the product but not in any of the pre-constructed Profile Sets.

Account numbers

An account number is the primary identifier for ownership of an account, whether a vendor account, a checking or brokerage account, or a loan account. An account number is used whether or not the identifier uses letters or numbers. Below are the profile Expressions Delphix uses to identify account numbers:

Expression Name	Domain	Expression Level	Expression
Account number	ACCOUNT_NO	Column	(?>(acc(oun\ n)?t)?_?(num(ber)?\ nbrjno)?)(?!\\w*(ID\\ type))

Physical addresses

Below are the profile Expressions Delphix uses to identify physical addresses:

Expression Name	Domain	Expression Level	Expression
Address	ADDRESS	Column	^(?:(?!postalcode\ city\ state\ country\ email\ (l\ ln\ lin\ line)?_?2{1}\ ID).)*addre?s?s?_?(?:(?!city\ state\ country\ email (l\ ln\ lin\ line)?_?2{1}\ ID).)*\$
Street Address	ADDRESS	Column	(?>(str(eet)?_?addre?s?s?\ street))(?!\\w*(ID\\ type))
Data - Address	ADDRESS	Data	(.[\s]+b(ou)?*l(e)?v(ar)?d[\s].) (.[\s]+st[.]?reet)?[\s].) (.[\s]+ave[.]?nue)?[\s].) (.[\s]+r(oa)?d[\s].) (.[\s]+l(a)?n(e)?[\s].) (.[\s]+cir(cle)?[\s].*)
Address Line2 - before	ADDRESS_LIN E2	Column	^(?:(?!email\ ID).)*(l\ ln\ lin\ line)?2{1}_?addre?s?s?(?:(?!email\ ID).)*\$

Expression Name	Domain	Expression Level	Expression
Address Line2 - after	ADDRESS_LINE2	Column	<code>^(?:(!email\ ID).)*address?_?(l\ ln\ lin\ line)?_?2{1}(?:(!email\ ID).)*\$</code>
Data - Address Line 2	ADDRESS_LINE2	Data	<code>(.*[\s]*ap(ar)?t(ment)?[\s]+.*) (.*[\s]*s(ui)?te[\s]+.*)\ (c(are)?[\s]*[\\\/]?o(f)?[\s]+.*)</code>

Beneficiary ID

Below are the profile Expressions Delphix uses to identify beneficiary IDs:

Expression Name	Domain	Expression Level	Expression
Beneficiary number	BENEFICIARY_NO	Column	<code>(?>(beneficiary)?_?(num(ber)? nbr\ no))(?!w*ID)1</code>
Beneficiary ID	BENEFICIARY_NO	Column	<code>(?>(beneficiary)?_?id)</code>

Biometrics

Below are the profile Expressions Delphix uses to biometric data:

Expression Name	Domain	Expression Level	Expression
Biometric	BIOMETRIC	Column	biometric

Certificate ID

Below are the profile Expressions Delphix uses to identify certificate IDs:

Expression Name	Domain	Expression Level	Expression
Certificate number	CERTIFICATE_NO	Column	(?>cert(ificate)?_(num(ber)?\ nbr\ no\ id))
Certificate ID	CERTIFICATE_NO	Column	(?>cert(ificate)?_?id)

City

Below are the profile Expressions Delphix uses to identify cities:

Expression Name	Domain	Expression Level	Expression
City	CITY	Column	ci?ty(?!w*ID)

Country

Below are the profile Expressions Delphix uses to identify countries:

Expression Name	Domain	Expression Level	Expression
Country	COUNTRY	Column	c(ou)?nty(?!w*ID)

Credit card

Below are the profile Expressions Delphix uses to identify credit cards:

Expression Name	Domain	Expression Level	Expression
Card number	CREDIT CARD	Column	(?>ca?rd_(num(ber)?\ nbr\ no)?)(?!w*ID)
Credit Card number	CREDIT CARD	Column	(?>cre?di?t_(ca?rd)?_(num(ber)?\ nbr\ no)?)(?!w*ID)

Expression Name	Domain	Expression Level	Expression
Data - Credit Card	CREDIT CARD	Data	^(?:3[47][0-9]{13} 4[0-9]{12}(?:[0-9]{3})?(?:[0-9]{3})?)\ (?:5[1-5][0-9]{2}\ 22[1-9]\ 22[3-9][0-9]\ 2[3-6][0-9]{2}\ 27[01][0-9]\ 2720)[0-9]{12}\ 6(?:011\ 5[0-9][0-9])[0-9]{2}\ 4[4-9][0-9]{3}\ 2212[6-9]\ 221[3-9][0-9]\ 22[2-8][0-9]{2}\ 229[0-1][0-9] 2292[0-5])[0-9]{10}(?:[0-9]{3})?)\ 3(?:0[0-5,9]\ 6[0-9])[0-9]{11}\ 3[89][0-9]{14}(?:[0-9]{1,3})?)\$

Customer number

Below are the profile Expressions Delphix uses to identify customer IDs:

Expression Name	Domain	Expression Level	Expression
Customer number	CUSTOMER_NUM	Column	(?>(cu?st(omer\ mr)?_?(num(ber)?\ nbr no?))(?!\\w*ID)

Date of birth

Below are the profile Expressions Delphix uses to identify dates of birth:

Expression Name	Domain	Expression Level	Expression
Birth Date	DOB	Column	(?>(bi?rth)_?(date?\ day\ dt))(?!\\w*ID)
Birth Date1	DOB	Column	(?>dob\ dtofb\ (day\ date?\ dt)_?(of)?_?(bi?rth))(?!\\w*ID)
Birth Date2	DOB	Column	(?>b_?(date?\ day))(?!\\w*ID)
Admission Date	DOB	Column	(?>(adm(it\ ission)?_?(date?\ day\ dt))(?!\\w*ID)

Expression Name	Domain	Expression Level	Expression
Treatment Date	DOB	Column	(?>(tr(ea)?t(ment)?_?(date?\ day dt))(?! \w*ID)
Discharge Date	DOB	Column	(?>(ds\ disc(h harge)?_?(date?\ day\ dt)) (?! \w*ID)

Driver license number

Below are the profile Expressions Delphix uses to identify driver license numbers:

Expression Name	Domain	Expression Level	Expression
Drivers License number	DRIVIN G_LC	Column	(?>(dri?v(e?rs?e?))_?(license li?c)?_? (num(ber)?\ nbr no?))(?! \w*ID)
Drivers License number1	DRIVIN G_LC	Column	(^license\$\ (license\ li?c)_?(num(ber)?\ nbr\ no))(?! \w*ID)

Email

Below are the profile Expressions Delphix uses to identify emails:

Expression Name	Domain	Expression Level	Expression
Email	EMAIL	Column	^(?(?!invalid).)*email(?! \w*ID)
Data - Email	EMAIL	Column	\b[A-Z0-9._%+-]+@[A-Z0-9.-]+\.[A-Z] {2,6}\b

First name

Below are the profile Expressions Delphix uses to identify first names:

Expression Name	Domain	Expression Level	Expression
First Name	FIRST_NAME	Column	(?>(fi?rst)?(na?me?)\ f_?name)(?! \w*ID)
Middle Name	FIRST_NAME	Column	(?>(mid(dle)?)?(na?me?)\ m_?name) (?!\\w*ID)

IP address

Below are the profile Expressions Delphix uses to IP addresses:

Expression Name	Domain	Expression Level	Expression
IP Address	IP ADDRESS	Column	(?>(ip_?address?s?s?)) (?!\\w*(ID\\ type))
Data - IP Address	IP ADDRESS	Data	\\b(?:(:?25[0-5]\\ 2[0-4][0-9]\\ 1[0-9][0-9]\\ [1-9]?[0-9])\\.) {3} (:?25[0-5]\\ 2[0-4][0-9]\\ 1[0-9][0-9]\\ [1-9]?[0-9])\\b

Last name

Below are the profile Expressions Delphix uses to identify last names:

Expression Name	Domain	Expression Level	Expression
Last Name	LAST_NAME	Column	^(?:(!portal\\ ID).)*((last)?(na?me?)\\ l_?name)(?:(!portalname\\ ID).)*\$

Plate number

Below are the profile Expressions Delphix uses to identify plate numbers:

Expression Name	Domain	Expression Level	Expression
License Plate	PLATE_NO	Column	<code>^(?:(!template ID type).)*(license\ li?c)?_?plate_?(num(ber)?\ nbr\ no)?(?:(!template\ ID\ type).)*\$</code>

PO Box numbers

Below are the profile Expressions Delphix uses to identify PO box numbers:

Expression Name	Domain	Expression Level	Expression
PO Box	PO_BOX	Column	<code>po_?box</code>
Data - PO Box	PO_BOX	Data	<code>po box\ p\.o\</code>

Precinct

Below are the profile Expressions Delphix uses to identify precincts:

Expression Name	Domain	Expression Level	Expression
Precinct	PRECINCT	Column	<code>(>?precinct\ prcncnt)(?!w*ID)</code>

Record number

Below are the profile Expressions Delphix uses to identify record numbers:

Expression Name	Domain	Expression Level	Expression
Record number	RECORD_NO	Column	<code>(?>rec(ord)?_?(num(ber)?\ nbr\ no))(?!w*(ID\ type))</code>

School name

Below are the profile Expressions Delphix uses to identify school names:

Expression Name	Domain	Expression Level	Expression
School Name	SCHOOL_NM	Column	(?>school_?na?me?)(?! \w*ID)

Security code

Below are the profile Expressions Delphix uses to identify security codes:

Expression Name	Domain	Expression Level	Expression
Security Code	SECURITY_CODE	Column	(?>se?cu?r(i?ty?)?_?co?de?)(?! \w*ID)

Serial number

Below are the profile Expressions Delphix uses to identify serial numbers:

Expression Name	Domain	Expression Level	Expression
Serial number	SERIAL_N M	Column	(?>(ser(ial)?)_?(num(ber)?\ nbr no)) (?!\\w*ID)

Signature

Below are the profile Expressions Delphix uses to identify signatures:

Expression Name	Domain	Expression Level	Expression
Signature	SIGNATURE	Column	signature(?!\w*(ID\ type))

Social security number

Below are the profile Expressions Delphix uses to social security numbers:

Expression Name	Domain	Expression Level	Expression
Social Security number	SSN	Column	ssn(?!\w*ID)

Expression Name	Domain	Expression Level	Expression
Data - SSN	SSN	Data	<code>\b(?:000)(?:!666)[0-8]\d{2}[-](?:!00)\d{2}[-](?:!0000)\d{4}\b</code>

Tax ID

Below are the profile Expressions Delphix uses to identify tax IDs:

Expression Name	Domain	Expression Level	Expression
Tax ID number	TAX_ID	Column	<code>tin\$ ^tin\ _tin\ tin_</code>
Tax ID Code or number	TAX_ID	Column	<code>(ta?x)?_(id(ent)?)?_?((co?de?)\ num(ber)?\ nbr\ no))?</code>

Telephone number

Below are the profile Expressions Delphix uses to identify telephone numbers:

Expression Name	Domain	Expression Level	Expression
Telephone or Contact number	TELEPHONE_NO	Column	<code>(?>((tele?)?phone)\ (co?nta?ct\ tel)_?(num(ber)?\ nbr\ no))(?!w*(ID\ type))</code>
Data - Phone number	TELEPHONE_NO	Data	<code>\(?:\b[0-9]{3}\)\?[-.]?[0-9]{3}[-.]?[0-9]{4}\b</code>
Fax number	TELEPHONE_NO	Data	<code>(?>fax_?(num(ber)?\ nbr\ no)?)(?!w*(ID\ type))</code>

Vin number

Below are the profile Expressions Delphix uses to identify vin numbers:

Expression Name	Domain	Expression Level	Expression
Vehicle	VIN_NO	Column	vehicle
VIN	VIN_NO	Column	vin\$ ^vin\ _vin\ vin_

Web address

Below are the profile Expressions Delphix uses to identify web addresses:

Expression Name	Domain	Expression Level	Expression
Web or URL Address	WEB	Column	(?>(url\ web_?adre?s?s?))(?!w*(ID\ type))
Data - Web Address	WEB	Data	\b(?:(:https?\ ftp\ file)://\ www\. \ ftp\.)[-AZ0-9+&-@#/%=~_\ \$?!:,.*][A-Z0-9+&-@#/%=~_\ \$\]

ZIP code

Below are the profile Expressions Delphix uses to identify zip codes:

Expression Name	Domain	Expression Level	Expression
zip or Postal Code	ZIP	Column	(?>(zip\ post(al)?_?((co?de?)?4?))(?!w*ID)
Data - Zip Code	ZIP	Data	1\b([0-9]{5})-([0-9]{4})\b

Configuring profile sets

A profile set defines the set of classifiers or expressions that will be used to identify sensitive information in the rule set when a profiling job is run. Refer to [Discovering Your Sensitive Data](#) for an overview of profile sets and related concepts.

To display, view, and manage the profile sets, click on the **Settings** tab and select **Profiler Sets** on the left-hand side of the page.

Home > Settings > Profile Sets

Settings Add Profile Set

Profile Sets

Algorithms
Domains
Profile Sets
Classifiers
Expressions
Roles
File Formats
JDBC Drivers

Displaying 1 to 4 of 4

Name	Description	Owner	Created	Action
Financial - Legacy		System	Feb 22, 2023 12:00 AM EST	...
HIPAA - Legacy		System	Feb 22, 2023 12:00 AM EST	...
Standard		System	Feb 22, 2023 4:24 PM EST	...
ASDD Standard		System	Feb 22, 2023 4:24 PM EST	...

The **Profiler Sets** screen displays each available profile set, along with an action button containing **view**, **edit**, **duplicate**, and **delete** options (assuming the user role allows those operations).

Creating and modifying profile sets

i Creating or modifying profile sets can also be done via the API and requires only two API calls. See [ASDD Profile Set Import and Export](#) for usage instructions.

The content of a profiler set depends on the profiler implementation with which it is intended to be used. Profile sets for the legacy profile contain search expressions and type expressions, while profile sets for the ASDD profiler contain classifiers.

To add a Profiler set, click **Add Profiler Set** at the top of the **Settings > Profile Sets** screen. The Add Profiler Set dialog appears:

Add Profile Set

Set Name

New Profile Set

Description

Example

ASDD Support ⓘ

Search Expressions

Type Expressions

Select Search Expressions ▼

Close

Submit

To add a profiler set for the legacy profiler

1. Click **Add Profile Set** at the top of the dialog window.
2. Enter a profiler **Set Name**. This name must be unique among all profile set names.
3. Optionally, enter a **Description** for this Profiler Set.
4. Leave the **ASDD Support** box unchecked.
5. Select the **Search Expressions** tab (it is selected by default).
6. Click the down arrow to the right of **Select Search Expression** to expand the tree. Click the box to the left of each search expression you wish to add to the profile set. It is not recommended that All Search Expressions be selected, as there is significant duplication among built-in search expressions.
7. Select the **Type Expressions** tab.
8. Click the down arrow to the right of **Select Type Expression** to expand the tree. Click the box to the left of each type expression you wish to add to the profile set.
9. When you are finished adding expressions, click **Submit** to save the new profile set.

Add Profile Set

Set Name

New Profile Set

Description

Example

 ASDD Support ⓘ

Classifiers

Select Classifiers ▼

Close

Submit

To add a profile set for the ASDD Profiler

1. Click **Add Profile Set** at the top of the dialog window.
2. Enter a profiler **Set Name**. This name must be unique among all profile set names.
3. Optionally, enter a **Description** for this Profiler Set.
4. Check the **ASDD Support** box.
5. Select the **Classifiers** tab (it is selected by default).
6. Click the down arrow to the right of **Select Classifiers** to expand the tree. Click the box to the left of each classifier you wish to add to the profile set.
7. When finished with adding classifiers, click **Submit** to save the new profile set.

To edit an existing profile set

Select the **Edit** option from the ellipsis actions menu to the right of the profile set name. Expand the tab corresponding to the type object you'd like to add or remove from the profile set, and check or uncheck the box to the left of each object to add/remove them from the profile set as desired. Click **submit** to save your change.

ⓘ ASDD Support
It is not possible to change the value of the **ASDD Support** setting for an existing profile set. A new profile set must be created.

To delete an existing profile set

Select the **Delete** option from the ellipsis actions menu to the right of the profile set name. You will be blocked from deleting a profile set if it is currently assigned to any jobs.

Managing domains

Overview

This article describes how to create and manage domains. Refer to [Discovering Your Sensitive Data](#) for an overview of domains and related concepts.

 Creating or modifying profile sets can also be done via the API and requires only two API calls. See the [ASDD Profile Set Import and Export](#) article for usage instructions.

Domains

Domains identify a specific type of sensitive data, along with the masking algorithm to use for that data. From the **Settings** tab, click **Domains** to the left, the list of domains will be displayed. From here, you can add, edit, or delete domains.

Delphix Continuous Compliance includes built-in domains and algorithms for many common types of sensitive data. Users can choose to select a different default masking algorithm for a domain, and/or create additional domains with their own default algorithms.

Additional created algorithms appear in the **Algorithms** drop-down menu. Because each domain has a single default masking algorithm, a distinct domain (along with recognition logic) must exist or be created for each distinct algorithm in order for the profiler to assign that algorithm in rule sets.

If the purpose for the environment where a profile job is run is set to **Tokenization/Re-Identify**, the tokenization algorithm associated with the domain will be assigned instead of the masking algorithm. Each domain referenced by any profile set used in tokenization environments should have a tokenization algorithm value defined.

The **Settings > Domains** tab is where to define domains, along with their default masking and tokenization algorithms.

Adding a new domain

1. At the top of the **Domains** tab, click **Add Domain**.
2. Enter the new **Domain Name**. The domain name specified will appear as a menu option on the **Inventory** screen elsewhere in the Delphix Masking Engine. Domain names must be unique.
3. Select a default **Masking Algorithm** for the new domain, and click Next.
4. Select a default **Tokenization Algorithm** for the new domain, if desired, and click Next.
5. Click **Save**.

Add Domain ✕

- Domain Name**
- Algorithm
- Tokenized Algorithm
- Summary

Domain Name

Specify details to identify this domain.

Cancel Back Next Save

- To delete any domain, click the **Actions** button to the far right of the domain name and select the option **Delete**.

Home > Settings > Domains

Settings

Add Domain

Domains

↻

Displaying 1 to 10 of 45

Name	Owner	Masking Method	Tokenization Method	Actions
ACCOUNT_NO	System	AccNoLookup		⋮
ACCOUNT_TK	System	NullValueLookup	AccountTK	⋮
ADDRESS	System	AddrLookup		⋮
ADDRESS_LINE2	System	AddrLine2Lookup		⋮
BANK_ACCOUNT_NO	System	dlpx-core:CM Numeric		⋮
BENEFICIARY_NO	System	dlpx-core:CM Alpha-Numeric		⋮
BIOMETRIC	System	NullValueLookup		⋮
CERTIFICATE_NO	System	dlpx-core:CM Alpha-Numeric		⋮
CITY	System	USCitiesLookup		⋮
COUNTRY	System	dlpx-core:Countries SL		⋮

- View
- Edit
- Delete

[Environments](#) | [Monitor](#) | [Settings](#) | [Admin](#) | [Audit](#)

[Support](#)

Note that updating the default masking or tokenization algorithm assigned to a domain only impacts algorithm assignments made by future profiling job executions, it does **not** have any immediate affect on algorithm assignments in existing rule sets.

Managing classifiers

Classifier instances define the logic that the ASDD profiler uses to identify sensitive information. For an overview of classifiers and related concepts, refer to [Discovering Your Sensitive Data](#). Each classifier instance is based on a **classifier framework** that implements the recognition logic.

An overview of the available frameworks is available in the [classifier concept](#) section. The [API Calls for Managing Classifiers](#) article describes how to use the API Client to retrieve a detailed description of all classifier frameworks on the system, along with their configuration schemas.

To view a list of all classifier instances available, click **Classifiers** under the **Settings** tab:

Home > Settings > Classifiers

Settings

Classifiers Add Classifier

Displaying 1 to 10 of 135

Name	Domain	Owner	Type	Action
School Name - Path	SCHOOL_NM	System	PATH	...
Serial Number - Path	SERIAL_NO	System	PATH	...
IP Address - Path	IP ADDRESS	System	PATH	...
Email Address - Path	EMAIL	System	PATH	...
Security Code - Path	SECURITY_CODE	System	PATH	...
Last Name - Path	LAST_NAME	System	PATH	...
Signature - Path	SIGNATURE	System	PATH	...
First Name - Path	FIRST_NAME	System	PATH	...
Certificate Number - Path	CERTIFICATE_NO	System	PATH	...
Address Line 2 - Path	ADDRESS_LINE2	System	PATH	...

Environments | Monitor | Settings | Admin | Audit Support

To add a classifier

1. Click the **Add Classifier** button at the upper right.

Add Classifier

Classifier Name _____ Domain **Select Domain** Classifier Framework **Regex**

Description (Optional) _____

Regular Expressions

Match Strength	Checksum Type	Case Sensitive	Partial Match	Action

Reject Strength (Optional) **100**

Add Regex Configuration

Profiling Regex _____

Data Clean Regex (Optional) _____

Checksum Type (Optional) **None**

Match Strength (Optional) **100**

Cancel Add

Cancel Save

2. Select a Domain from the **Domain** dropdown.
 - Domains are used by profiling jobs to determine the masking algorithm to apply to the sensitive data. When a classifier is matched, the profiling job will associate the specified domain with the sensitive data. The Masking Engine comes out of the box with over 30 pre-defined domains. Domains can be added, edited, and deleted from the **Settings Domains** screen. For more information on domains, refer to the [Managing Domains](#) article.
3. Enter the following information for the classifier:
 - a. Classifier **name** to create.
 - b. **Description** for the classifier (optional).
4. Select the **Classifier Framework** to create the classifier with (i.e. Data Type, Regex, Path or List).
5. The bottom part of the form will change based on the classifier framework selected.
6. To Add a Regex Framework:
 - a. Fill out the **Profiling Regex** field (mandatory), then click **add**. For the rest of the fields, if left unfilled/unselected, the default value will be selected and added to the table.
 - b. Click **save** to create the classifier. Users can add multiple configurations of the same framework type to a classifier.
7. It also allows in-place editing/deleting of configurations.
 - a. To edit/delete, click the action button next to the configuration to edit/delete, then select the corresponding option.
8. Clicking **edit** will take all the information from the row to the form on the right side, where it can then be modified.

Add Classifier

Classifier Name: Postcode - Regex Domain: ZIP Classifier Framework: Regex

Description (Optional):

Regular Expressions

Match Strength	Checksum Type	Case Sensitive	Partial Match	Action
Profiling Regex : ^[0-9]{5}\$				
20%	NONE	No	Yes	...
Profiling Regex : ^[0-9]{5}(?:-[0-9]{4})\$				
70%	NONE	No	Yes	...

Add Regex Configuration

Profiling Regex: //

Data Clean Regex (Optional): //

Checksum Type (Optional): None

Match Strength (Optional): 100

Cancel Add

Reject Strength (Optional): 100

Cancel Save

To edit a classifier

Click the ... indicator in the **Action** column, to the right of the Classifier on the classifier grid, and choose **Edit**.

To delete a classifier

Click the ... indicator in the **Action** column, to the right of the Classifier on the classifier grid, and choose **Delete**.

Configuration considerations for classifiers

Classifier design is more complex than search or type expressions because classifiers offer more flexibility in matching logic and configuration around match strength. Classifiers contain more configuration, typically encompassing all logic of the framework's type for a particular domain. For example, a legacy profile set might have three different column-level search expressions, but these would all be consolidated into a single PATH type classifier. Classifiers also add the notion of rejection strength, which allows the profiling logic to eliminate domains from consideration earlier in the profiling process.

Strength values

Classifiers use a scale of -1.0 to 1.0 to represent match and reject strength, which correspond to the -100 to 100 confidence values used in the UI. The product currently does not display confidence values for non-matches, so only values between 0 and 100 are typically visible.

A value of -1.0 indicates absolute rejection, and the domain in question is immediately eliminated from the set of possible matches. Similarly, a value of 1.0 indicates absolute confirmation, and the domain is assigned as a match without checking any other classifiers for that domain.

When multiple classifiers produce match or reject strength values, those results may be combined to get a final confidence. If the results conflict, with opposite signs, the result with the highest absolute value takes precedence. If the results have the same sign, the final result for that domain is a stronger match. The exact values and formula applied are under development and may change in the future. Currently, only the strongest column level result and strongest data level result are combined in this fashion.

Examples

1. A column named "ssn_present" matches a PATH classifier for the SSN domain with a match strength of 0.67. However, the column is boolean type and does not match the TYPE classifier for the SSN domain, which returns a -1.0 result. The verdict is -1.0 and the SSN domain is not assigned.
2. A column named "passport_no" contains 9 digit numeric values, which match the REGEX classifiers for both the SSN and PASSPORT_NO domains. Both REGEX classifiers return a confidence of 0.5 for this match. However, while the PATH classifier for the PASSPORT_NO domain matches and returns a match strength of 0.67, the PATH classifier for the SSN domain does not match, returning 0. The final confidence values are PASSPORT_NO at 0.84, and SSN at 0.5, so the PASSPORT_NO domain is the best match and the PASSPORT_NO domain and associated masking algorithm are assigned to the column.



Default assignment threshold

The ASDD profiler has a default minimum confidence value of 1, which means that any positive match, no matter how weak, will trigger an assignment. This is significantly different from the legacy profiler, which by default requires an 80% match for data level expressions. The application setting ASDD/DefaultAssignmentThreshold controls this value. For more details, refer to this section.

Choosing values for match strength

The match strength value (typically called **matchStrength** in the classifier configuration) reflects how confident the classifier is that a particular data element exclusively matches the associated domain. A match strength of 0.01 indicates that the data element may belong to the domain, but might also belong to any number of other domains or not be sensitive at all, while a value of 1.0 reflects absolute certainty that this data matches this domain and no other domain. A value of 0 provides no information.

Not all classifiers have a match strength greater than 0. One example of this is TYPE classifiers, which typically have a high reject strength, but 0 match strength (since it is impossible to match any of the built-in domains based on the data type of a column alone).

PATH classifiers built-in to the product typically have a match strength of 0.67, so in order for a REGEX or LIST classifier to override a PATH result, that classifier's match strength or reject strength would have to be higher than this value. This can help eliminate false positive results from the PATH classifiers, but be wary of the next recommendation before setting match strength to a high value.

When choosing match strength for REGEX classifiers, consider whether the pattern is unique to the type of sensitive data being detected. If it is not, it is safer to give a relatively low match strength in the range of 0.1 to 0.5, so that PATH level results can contribute information. Consider this example of REGEX detection of US Social Security numbers. These might be stored as a string value with a more distinct pattern like "001-23-4567", or simply as a 9-digit number "001234567". A 9-digit number might be any number of other numeric identifiers, like account number, passport number, a row identifier for rows in another table, etc. so the match strength for the [0-9]{9} regex should be quite low. The distinct text pattern with dashes has a much higher match strength since it is unlikely to be any other kind of information.

Here are some additional tips for choosing match strength:

- Use a higher match strength for patterns that are more likely to be unique to the type of sensitive data being detected.
- Use a lower match strength for patterns that are less likely to be unique to the type of sensitive data being detected.
- Use a match strength of 0 for patterns that are not unique to the type of sensitive data being detected.

Choosing values for reject strength

The reject strength (typically called **rejectStrength** in the classifier configuration) value reflects how likely it is that a value matches the classifier's domain when the classifier does not match. If you are certain that your classifier configurations will match every possible value for the domain, the reject strength should be set to 1.0; however, this degree of certainty is rare. Similar to match strength, not all classifiers provide any rejection capability. This is true of PATH classifiers, for example, as we cannot rely on an unknown database schema to use predictable or human-readable names for columns.

The reject strength for classifiers applies any time there is no match. For example, if a REGEX classifier contains 4 regular expressions, each expression would be tested against the column data value, and if none match, the reject strength defines the result. For this reason, it can be useful to add a pattern that matches quite broadly, even if it's not particularly selective for the domain in question, with a low match strength. This prevents a full rejection for values that might match this classifier's domain as well as one or more other domains.

For LIST and REGEX classifiers where the set of patterns or list values is known to be only a subset of possible values for the domain, reject strength should be below 0.5 to allow column level matches to take precedence, even if none of the data values match. For example, the value lists built-in for first and last name LIST classifier only contain English values, and names might be in other languages. These classifiers have "reject strength" set relatively low, to prevent the LIST classifiers from overriding the PATH classifier match if, for example, the column contains only Japanese names.

Here are some additional tips for choosing reject strength:

- Use a higher reject strength for classifiers that are more likely to match all values in the domain.
- Use a lower reject strength for classifiers that are less likely to match all values in the domain.
- Use a reject strength of 0 for classifiers that are not designed to match all values in the domain.

Regex configuration

The PATH and REGEX classifier types consume regular expressions using Java 8 regex syntax and matching logic. These classifiers have additional configuration options to control whether these patterns should match the entire input, and whether they are case-sensitive. For this reason, avoid using regex constructs such as "**^(pattern)\$**" for these purposes.

Type classifiers

The TYPE classifier framework uses the same four types as Type Expressions, as described in the [Managing expressions](#) section. However, the type-matching system is more versatile and provides better type identification across all database variants. This framework also supports the identification of auto-incrementing columns. The property **matchAutoIncrement** defines whether auto-incrementing columns should be considered for profiling (**true**) or completely skipped (**false**) for the domain. The default value is **false**.

Managing expressions

The **search** and **type** expressions define logic used by the legacy profiler to identify sensitive information. Refer to [Discovering Your Sensitive Data](#) for an overview of expressions and related concepts.

Profile expressions

The **column** and **data** level expressions use regex text patterns on column meta-data and the data within the column, respectively, to identify sensitive data.

 Column and data level expressions are case insensitive.

Type expressions can limit matches with column-level expressions by data type. A Type Expression consists of a user-chosen name, a data type, an optional minimum field length, and a domain to which the constraint applies. The supported data types are String, Number, Date, and Binary. Each type represents a number of native datatypes in the database.

- For **String** type, all character types supported by the database such as VARCHAR, NVARCHAR, CLOB, and NCLOB are considered String types for profiling. The minLength parameter considers the length specification of the column type, which may be characters or bytes. For example, Oracle supports VARCHAR2 fields measuring in either characters or in bytes. A VARCHAR2(20) column can hold 20 characters, whereas a VARCHAR2(20 BYTE) column can hold 20 bytes, which may be fewer than 20 characters if multibyte characters are present. A type expression with a minLength of 20 will match to both.
- For **Number** type, all numeric types are considered Number types by the profiling logic, including INTEGER, FLOAT, BIG_INTEGER, etc. The minLength parameter considers the number of base-10 digits supported by the type. For floating-point values, minLength refers to the integral part of the number.
- For **Date** type, the Date type includes all calendar date and date/time types, such as DATE and LOCAL_DATE_TIME types. The minLength parameter is not permitted for Date Type Expressions.
- For **Binary** type, the Binary type includes large object types such as BLOB and BINARY. The minLength parameter considers the maximum storage size of the column in bytes.

If there is more than one Type Expression assigned to a domain, then a column will match for the domain if the regular expression matches, and at least one of the type expressions match. For example, dates of birth are often stored in string types instead of dates, so you might have a string type expression and a date type expression assigned to the Date of Birth domain to allow columns of either type to match. Two Type Expressions of the same type cannot be assigned to the same domain in the same profile set. If there are no Type Expressions assigned to a domain, then the profile expression alone will determine matching without regard to data type.

Like Profile Expressions, Profile Type Expressions must be part of a profile set to be effective. Profile Type Expressions have no effect on Data Level Profiling.

 Profile Type Expressions are only supported for database profiling. They have no effect on profiling of file data.

 As of version 9.0, only Oracle and MSSQL Server are fully supported. On other platforms, Type Expressions may result in unexpected matches.

Managing expressions

In order to manage search and Type Expressions, select **Expressions** using the navigation panel on the left-hand side of the **Settings** tab.

Home > Settings > Profiler

Settings Profiler

Algorithms
Domains
Profile Sets
Classifiers
Expressions
Roles
File Formats
JDBC Drivers

Search Expressions Type Expressions Add Search Expression

C Displaying 1 to 11 of 96

Domain	Expression Name	Owner	Level	Action
DOB	Birth Date	System	Column	...
ADDRESS	Address	System	Column	...
ADDRESS_LINE2	Address Line 2 - after	System	Column	...
ACCOUNT_NO	Account Number	System	Column	...
DOB	Birth Date1	System	Column	...
CUSTOMER_NO	Customer Number	System	Column	...
CREDIT CARD	Credit Card Number	System	Column	...
CREDIT CARD	Card Number	System	Column	...
DRIVING_LC	Drivers License Number	System	Column	...
DRIVING_LC	Drivers License Number1	System	Column	...

This panel has **Search Expression** and **Type Expression** tabs that select which type of expression is visible.

To add a search expression

1. Click the **Search Expression** tab near the top of the **Expressions** screen.
2. Click the **Add Search Expression** button at the upper right.
3. Select a Domain from the **Domain** dropdown.
 - Domains are used by Profiling jobs to determine the masking algorithm to apply to your sensitive data. When an Expression is matched, the Profiling job will associate the specified Domain to the sensitive data. The Masking Engine comes out of the box with over 30 pre-defined Domains. Domains can be added, edited, and deleted from the **Settings Domains** screen.
4. Enter the following information for the Expression:
 - **Expression Name:** The name used to select this expression as part of a Profiler Set.
5. Select an **Expression Level** for the Expression:
 - **Column Level:** To identify sensitive data based on column names.
 - **Data Level:** To identify sensitive data based on data values, not column names.
6. Enter the **Expression Text:** The regular expression used to identify sensitive data.

Add Search Expression

Expression Name

Domain
Select Domain ▼

Expression Level
Select Expression Level ▼

Regular Expression

To add a type of expression

- Click the **Type Expression** tab near the top of the **Expressions** screen.
- Click the **Add Type Expression** button at the upper right.
- Set a value for **Expression Name** and select a **Domain** as you would for a search expression.
- Select **Constraints** (Data Type) for the expression: String, Numeric, Binary, Date.
- Set a **Minimum Column Length** for the data type if desired.

Note: Length constraints are not applied to large object types such as CLOBs and BLOBs.

For example, to ensure that column-level profiling only identifies a column with the FIRST_NAME domain, if the column is a string type and has a capacity of at least 5 characters, add the type constraint shown below.

Add Type Expression

Expression Name	FirstNameType
Domain	FIRST_NAME
Expression Level	Type Level
Constraints	STRING
Minimum Column Length	5

- When you are finished, click **Save**.

To edit an expression

Click the ... indicator in the **Action** column to the right of the Expression and choose **Edit**.

To delete an expression

Click the ... indicator in the Action column to the right of the Expression and choose Delete. Deletion will be blocked if the expression is currently assigned to one or more profile sets.

Searching and filtering expressions

The `profile-expressions/search` and `profile-type-expressions/search` endpoints allow for searching and filtering of profile expressions and profile type expressions. More information on syntax can be found at [API Calls for Searching and Filtering](#).

ASDD profile set import and export

The ASDD Asynchronous Discovery set upgrade allows users to seamlessly create and update an ASDD profile set and its associated domains and classifiers. For usage instructions, see [API Calls for ASDD Profile Set Import and Export](#).

ASDD profile set import

ASDD profile set import facilitates a quicker and simpler option for the creation or update of the following masking objects:

1. A profile set
 - A new profile set can be created.
 - An existing profile set can be updated **if there are no active profiling jobs using the specified profile set.**
2. Domains
 - New domains can be created along with their assigned masking and tokenization algorithms.
 - Existing domains can have their assigned masking and tokenization algorithms updated.
3. Classifiers
 - New classifiers can be created.
 - Existing classifiers' configuration can be modified **if there are no active profiling jobs using a profile set that includes the specified classifiers.**

The profile set configuration is exported in the form of a dynamically generated JSON file, formatted just like the [ASDD Standard profile set JSON](#), including any files used by the classifiers that are a part of the profile set, all packaged in a single zip file.

Limitations

ASDD profile set import does not support importing multiple profile sets at a time; it only allows for a single profile set configuration to be installed at a time.

ASDD profile set export

Like the import function, the ASDD profile set export allows for an easy option to export a given profile set configuration in a zip file. This can be used to easily import the profile set onto another engine, or modifications can be made to the profile set JSON and then imported to the same engine to make updates to the given profile set.

Creating a profiling job

This section describes how users can create a Profiling job. You can create Profiling jobs for databases, XML, mainframe files, delimited files, and fixed-width file rule sets. It is not currently possible to profile XML or JSON documents stored in database columns.

When a profiling job runs, it applies all of the recognition logic specified in the profile set to each data element present in the rule set. The behavior of the profiler is also influenced by several application settings, refer to the **Profile group settings** section of [this article](#).

The Profiler assigns each sensitive data element to a domain, with each domain having a default masking algorithm. Then, in the inventory, masking algorithms can be manually updated as needed to establish the masking rulesets for your data sources.



Column and Field Priority

If you wish to prevent the profiler from updating the domain and algorithm assignments for a particular column or file field, set the Priority value for the column or field to USER.

Profiling Jobs are grouped within environments on the **Environment Overview** page along with all masking jobs. In order to navigate to the **Overview** screen, click on an environment and the **Overview** tab should automatically display.

The screenshot displays the DELPHIX MASKING web application interface. The top navigation bar includes 'Job Wizard' and 'admin'. The main navigation menu has 'Environments', 'Monitor', 'Settings', 'Admin', and 'Audit'. The 'Environments' section is active, showing tabs for 'Overview', 'Connector', 'Rule Set', and 'Inventory'. The breadcrumb trail is 'Home > Environments > test1'. The environment name 'test1' is prominently displayed. To the right of the name are three buttons: 'Export', 'Profile', and 'Mask'. Below this is a table with the following data:

Environment	
Name	test1
Purpose	Mask
Application Name	test1
Approval workflow	Disabled

At the bottom, there is a table header with columns: Job ID, Name, Rule Set, Completed, Status, Action, Edit, and Delete. The footer includes navigation links: 'Environments | Monitor | Settings | Admin | Audit' and the DELPHIX logo.

Creating a new profiling job

To create a new Profiling job:

1. Click the **Profile** button on the upper side of the page.
2. The **Create Profiling Job** window appears.

Create Profile Job

Job Name

Feedback Size

Target: test1

Multi Tenant

Rule Set

No. of Streams

Min Memory **Max Memory**

Multiple Profiler Expression Check

Profile Sets

Comments

Email

3. You will be prompted for the following information:

- **Job Name:** A free-form name for the job you are creating. Must be unique.
- **Multi Tenant:** Check the box if the job is for a multi-tenant database. This option allows existing rulesets to be re-used to mask identical schemas via different connectors. The connector is selected at job execution time.
- **Rule Set:** Select the rule set that this job will profile.
- **No. of Streams:** The number of parallel streams to use when running the jobs. For example, you can select two streams to profile two tables in the ruleset concurrently in the job instead of one table at a time.
- **Min Memory (MB)** (optional): Minimum amount of memory to allocate for the job, in megabytes.
- **Max Memory (MB)** (optional): Maximum amount of memory to allocate for the job, in megabytes. When an ASDD profile set is selected, the max memory for the job must be at least 1024MB for each stream. For example, if No. of Streams is 4, this value would need to be 4096 or higher.
- **Feedback Size** (optional): The number of rows to process before writing a message to the logs. Set this parameter to the appropriate level of detail required for monitoring your job. For example, if you set this number significantly higher than the actual number of rows in a job, the progress for that job will only show 0 or 100%.
- **Multiple Profiler Expression Check:** By default, the profiler stops testing Profiler Expressions on a column or data value after the first expression matches. Check this box if the job should check all Profiler Expressions. If multiple Profiler Expressions match, the Profiler report will indicate multiple matches and the algorithm specified by the `DefaultMultiPhiAlgorithm` application setting will be assigned. This setting applies to both the legacy and ASDD profilers.

- **Profile Sets:** The name of the Profile Set to use. A Profile Set is a set of Profile Expressions (for example, a set of financial expressions) or classifiers. The profile set selected determines whether the legacy or ASDD profiler will run. If the current data source is not supported by the ASDD profiler, selecting an ASDD profile set will result in an error and another profile set must be selected. Refer to [this section](#) for information regarding which connectors are supported by ASDD.
 - **Comments** (optional): Add comments related to this job.
 - **Email** (optional): Add e-mail address(es) to which to send status messages. Separate addresses with a comma (,).
4. When you are finished, click **Save**.

Running a profiling job

This section describes how users can run a profiling job from the **Environment Overview** screen.

The screenshot shows the DELPHIX MASKING interface. At the top, there is a navigation bar with 'Environments', 'Monitor', 'Settings', 'Admin', and 'Audit'. A 'Job Wizard' button and a user dropdown 'admin' are on the right. Below the navigation bar, there are tabs for 'Overview', 'Connector', 'Rule Set', and 'Inventory'. The 'Overview' tab is selected. The breadcrumb path is 'Home > Environments > test1'. The environment name 'test1' is displayed prominently. To the right of the name are three buttons: 'Export', 'Profile', and 'Mask'. Below this is an 'Environment' details section with the following information:

Name	test1
Purpose	Mask
Application Name	test1
Approval workflow	Disabled

Below the details is a table of jobs:

Job ID	Name	Rule Set	Completed	Status	Action	Edit	Delete
1	ProfileJob	fileconnector	...	Created			

At the bottom of the screen, there is a navigation bar with 'Environments | Monitor | Settings | Admin | Audit' and the DELPHIX logo.

To run or rerun a job from the **Environment Overview** screen:

- Click the **Run** icon (play icon) in the Action column for the desired job.
- The **Run** icon changes to a **Stop** icon while the job is running.
- When the job is complete, the **Status** changes.

To stop a running job from the **Environment Overview** screen:

1. Locate the job you want to stop.
2. In the job's **Action** column, click the **Stop** icon.
3. A popup appears asking, "Are you sure you want to stop job?" Click **OK**.

When the job has been stopped, its status changes.

Reporting profiling results

This section describes the different ways of sharing/exploring the results of a profiling job.

Monitor page

After a Job has been started from the Environment **Overview** screen, clicking on the Job Name will result in the display of the profiling job from the **Monitor** tab. Clicking on the **Results** tab in the middle of the screen after the job has been completed will display the sensitive data findings on a table-column by table-column or file-field by file-field basis.

Home > Monitor > [Result Monitor](#)


DATABASE


SUCCESS

0
Jobs Running

oracle

<p>Job Type Profile</p> <p>Environment testenv</p> <p>Job ID</p> <p>Execution ID 1</p> <p>Connection Type table</p> <p>Source / Target - / BLACKBOX</p> <p>Profiling Report</p> <p>1_1.log Log</p> <p>Execution Logs</p>	<ul style="list-style-type: none">  Initializing  Collecting Configurations  Preparing  Starting  Profiling Completed 	<p>Start Time 14:33:21</p> <p>Previous Run Time</p> <p>Total # of Tables 1</p> <p>Tables Profiled 1</p> <p>Tables to be Profiled 0</p> <p>Total Time Taken (HH:mm:ss) 00:00:01</p> <p>Streams 1</p>
--	---	--

[Completed](#) [Processing](#) [Waiting](#) **Results**

Profiler Results

8
Sensitive

1
Total Tables

Table	Column	Domain	Algorithm
PROFILE_TEST	DOB	DOB	DateShiftDiscrete
PROFILE_TEST	ADDRESS	ADDRESS	AddrLookup
PROFILE_TEST	ZIPCODE	ZIP	RepeatFirstDigit
PROFILE_TEST	PHONE_NO	TELEPHONE_NO	d1px-core:Phone Unique
PROFILE_TEST	LAST_NAME	FULL_NAME	d1px-core:FullName
PROFILE_TEST	EMAIL	EMAIL	EmailLookup
PROFILE_TEST	FIRST_NAME	FIRST_NAME	FirstNameLookup
PROFILE_TEST	SSN	SSN	NullValueLookup

PDF report

To retrieve a PDF report from the **Results** tab, click on the **Profiling Report** link near the top of the page.

D E L P H I X

Profiling Report

Job Profile		Job Status	
Name	oracle	Current Status	SUCCEEDED
Type :	Profiling	Start Time :	2022-08-16 21:33:21,646
Environment :	testenv	End Time :	2022-08-16 21:33:23,369
Schema :	BLACKBOX		

Table	Column	Domain	Algorithm	Data Type	Confidence	Auto ID
PROFILE_TEST	CITY	CITY	NullValueLookup	VARCHAR2		false
PROFILE_TEST	COUNTRY			VARCHAR2		false
PROFILE_TEST	DOB	DOB	DateShiftDiscrete	DATE	100	true
PROFILE_TEST	ADDRESS	ADDRESS	AddrLookup	VARCHAR2	100	true
PROFILE_TEST	ZIPCODE	ZIP	RepeatFirstDigit	VARCHAR2	100	true
PROFILE_TEST	PHONE_NO	TELEPHONE_NO	dlpx-core:Phone Unique	VARCHAR2	100	true
PROFILE_TEST	LAST_NAME	FULL_NAME	dlpx-core:FullName	VARCHAR2	100	true
PROFILE_TEST	EMAIL	EMAIL	EmailLookup	VARCHAR2	100	true
PROFILE_TEST	FIRST_NAME	FIRST_NAME	FirstNameLookup	VARCHAR2	100	true
PROFILE_TEST	SSN	SSN	NullValueLookup	VARCHAR2	100	true

Inventory page

Alternatively, after a job completes successfully, the profiling results can be displayed through the **Inventory** screen. The inventory differs by connection type as shown below.

Database inventory

Profiling results can be determined by examining the assigned **Algorithm** for the table(s) in the Rule Set.

Overview Connector Rule Set **Inventory**

Home > Environments > testenv > Inventory > oracle

oracle

Import Export

Filter By: All Fields Masked Fields Auto User

Column	Data Type	Algorithm	Edit
ADDRESS	VARCHAR2 (256)	ADDRESS LINE SL	
CITY	VARCHAR2 (256)	NULL SL	
COUNTRY	VARCHAR2 (256)		
CREDIT_CARD	VARCHAR2 (30)		
DOB	DATE (7)	DATE SHIFT(DISCRETE)	
EMAIL	VARCHAR2 (256)	EMAIL SL	
FIRST_NAME	VARCHAR2 (256)	FIRST NAME SL	
LAST_NAME	VARCHAR2 (256)	dlpx-core:FullName	
PHONE_NO	VARCHAR2 (256)	dlpx-core:Phone Unique	
SSN	VARCHAR2 (256)	NULL SL	
ZIPCODE	VARCHAR2 (20)	ZIP+4	

Select Rule Set

oracle

Filter Contents

Search By Name

Search Alphabetically

PROFILE_TEST

Contents

PROFILE_TEST

File inventory

Profiling results can be determined by examining the assigned **Domain** and **Algorithm** for the files(s) in the Rule Set.

Overview Connector Rule Set **Inventory**

Home > Environments > testenv > Inventory > delimited

delimited

Filter By: All Fields Masked Fields Auto User

Import Export Record Types Define Fields

Select Rule Set

delimited

Filter Contents

Search By Name

Search Alphabetically

Formats

ff_Company_Emp_fi...

File

Company_Emp_file.txt

Record Type: All Records

Type	ID	Position	Method	Domain	Algorithm	Edit	Delete
ASCII	Address	5	Mask	ADDRESS	ADDRESS LINE SL		
ASCII	City	6	Mask	CITY	NULL SL		
ASCII	Country	9					
ASCII	Description	11					
ASCII	EmployeeID	2					
ASCII	FirstName	3	Mask	FIRST_NAME	FIRST NAME SL		
ASCII	LastName	4	Mask	FULL_NAME	dIpx-core:FullName		
ASCII	PhoneNumbers	10					
ASCII	State	7					
ASCII	TableSeq	1					
ASCII	ZipCode	8	Mask	ZIP	ZIP+4		

Mainframe Inventory

Profiling results can be determined by examining the assigned **Domain** and **Algorithm** for the files(s) in the Rule Set.

Home > Environments > testenv > Inventory > vsam
vsam

Filter By: All Fields Masked Fields Redefines Auto User

Select Rule Set

vsam

Filter Contents

Search By Name

Search Alphabetically

Formats

VSAM_Demo.cbl

File

VSAM_Demo.dat

VSAM_Demo.cbl

- CS-CUSTOMER-RECORD**
 - CUST-TYPE** EDIT
 - PERSON-DET** REDEFINED
 - PERSON-FIRSTNAME** MASKED
Domain: FIRST_NAME Algorithm: FIRST NAME SL
 - PERSON-LASTNAME** MASKED
Domain: FULL_NAME Algorithm: dlpx-core:FullName
 - PERSON-ADDRESS1** MASKED
Domain: ADDRESS Algorithm: ADDRESS LINE SL
 - PERSON-CITY** EDIT
 - PERSON-STATE** EDIT
 - PERSON-ZIP** MASKED
Domain: ZIP Algorithm: ZIP+4
 - PERSON-SSN** MASKED
Domain: SSN Algorithm: NULL SL
 - COMP-DET** REDEF
 - COMP-ENTITYNM** MASKED
Domain: FULL_NAME Algorithm: dlpx-core:FullName
 - COMP-ADDRESS1** MASKED
Domain: ADDRESS Algorithm: ADDRESS LINE SL
 - COMP-CITY** EDIT
 - COMP-STATE** EDIT
 - COMP-ZIP** MASKED
Domain: ZIP Algorithm: ZIP+4
 - COMP-PHONE** MASKED
Domain: TELEPHONE_NO Algorithm: dlpx-core:Phone Unique

CSV

To get a spreadsheet capturing the profiling results for the inventory, click on **Export** near the top of the page and a CSV file will be created.

Environment Name	Rule Set	Table Name	Type	Parent Column Name	Column Name	Data Type	Domain	Algorithm	Is Masked	ID Method	Row Type	Date Format	Notes	Multi-Column Logical Field	Group Number
testenv	oracle	PROFILE_TEST	-	-	DOB	DATE (7)	DOB	DATE SHIFT(DISCRETE)	TRUE	Auto	All Row	yyyy-MM-dd	-	-	-
testenv	oracle	PROFILE_TEST	-	-	LAST_NAME	VARCHAR2 (256)	FULL_NAME	dlpx-core:FullName	TRUE	Auto	All Row	-	-	-	-
testenv	oracle	PROFILE_TEST	-	-	SSN	VARCHAR2 (256)	SSN	NULL SL	TRUE	Auto	All Row	-	-	-	-
testenv	oracle	PROFILE_TEST	-	-	FIRST_NAME	VARCHAR2 (256)	FIRST_NAME	FIRST NAME SL	TRUE	Auto	All Row	-	-	-	-
testenv	oracle	PROFILE_TEST	-	-	EMAIL	VARCHAR2 (256)	EMAIL	EMAIL SL	TRUE	Auto	All Row	-	-	-	-
testenv	oracle	PROFILE_TEST	-	-	CREDIT_CARD	VARCHAR2 (30)	-	-	FALSE	Auto	All Row	-	-	-	-
testenv	oracle	PROFILE_TEST	-	-	PHONE_NO	VARCHAR2 (256)	TELEPHONE_NO	dlpx-core:Phone Unique	TRUE	Auto	All Row	-	-	-	-
testenv	oracle	PROFILE_TEST	-	-	CITY	VARCHAR2 (256)	CITY	NULL SL	TRUE	User	All Row	-	-	-	-
testenv	oracle	PROFILE_TEST	-	-	COUNTRY	VARCHAR2 (256)	-	-	FALSE	User	All Row	-	-	-	-
testenv	oracle	PROFILE_TEST	-	-	ZIPCODE	VARCHAR2 (20)	ZIP	ZIP+4	TRUE	Auto	All Row	-	-	-	-
testenv	oracle	PROFILE_TEST	-	-	ADDRESS	VARCHAR2 (256)	ADDRESS	ADDRESS LINE SL	TRUE	Auto	All Row	-	-	-	-

The spreadsheet can then be shared and manually modified to correct the sensitive data findings by:

1. Changing the **Is Masked**, **Algorithm**, and/or **Domains** fields for the respective Table/Column or File/Field in the CSV file accordingly.

2. Importing the modified spreadsheet by clicking on **Import** near the top of the **Inventory** screen and specifying the modified CSV file name.

API endpoint

Using the API endpoint `/profileResultDatabase/{executionId}`, profiling results can be retrieved by providing the executionId. This method is only for database connections and will not work with other connection types. Results will be returned in JSON format.

```
{
  "_pageInfo": {
    "numberOnPage": 4,
    "total": 4
  },
  "responseList": [
    {
      "columnMetadataId": 1,
      "columnName": "CITY",
      "tableName": "PROFILE_TEST",
      "domainName": "CITY",
      "algorithmName": "NullValueLookup",
      "dataType": "VARCHAR2",
      "isProfilerWritable": false
    },
    {
      "columnMetadataId": 2,
      "columnName": "COUNTRY",
      "tableName": "PROFILE_TEST",
      "dataType": "VARCHAR2",
      "isProfilerWritable": false
    },
    {
      "columnMetadataId": 3,
      "columnName": "DOB",
      "tableName": "PROFILE_TEST",
      "domainName": "DOB",
      "algorithmName": "DateShiftDiscrete",
      "dataType": "DATE",
      "confidence": 100,
      "isProfilerWritable": true
    },
    {
      "columnMetadataId": 4,
      "columnName": "ADDRESS",
      "tableName": "PROFILE_TEST",
      "domainName": "ADDRESS",
      "algorithmName": "AddrLookup",
      "dataType": "VARCHAR2",
      "confidence": 100,
      "isProfilerWritable": true
    }
  ]
}
```

}

ASDD features and support

The ASDD profiler was introduced in Continuous Compliance version 9.0, and represents the future direction for sensitive data discovery. It offers a number of advantages as compared to the legacy profiler, but currently has some limitations as well.

The introduction of the ASDD profiler does not make any changes to the legacy profiler. Existing profiling jobs should continue to function as they have in the past.

ASDD features

- The ASDD profiler uses classifiers rather than search and type expressions. Classifiers support more features and configuration options than expressions.
 - The LIST classifier framework is new has no equivalent functionality in the legacy profiler.
 - The TYPE classifier framework uses standard Java SQLType values to identify data types, which should provide broad support for type detection across all database variants.
 - The PATH classifier supports exact matching and can be configured to consider table name in addition to column name when matching.
 - The REGEX classifier supports detection of LUHN check digits for data level recognition of credit card numbers.
- The ASDD profiler provides better matching when the number of rows in a table is less than the target number of rows for profiling, and in general provides more nuanced confidence value in profiling results.
- The ASDD profiler attempts to retrieve more data values when a large fraction data values for a column are null or empty. The threshold to trigger an additional query is controlled by the [application setting ASDD/DefaultNullFilterThreshold](#).
- The ASDD profiler supports statistical sampling for Oracle and SQL Server databases, so that the data sampled will better reflect the full range of values for each column across the entire table.



Sample Percentage

When data sampling is employed, the sample percentage is always set to 1% - if this percentage does not yield enough rows, the query is performed again without sampling.

- The [ASDD Standard](#) profile set contains data level logic by default, allowing some columns containing sensitive information to be identified even if the column names are not meaningful.
 - New or improved REGEX classifiers for Zip Code and Email Address domains.
 - New LIST classifiers are present for First and Last Name, US City, US State and Country domains.
- Classifiers and profile sets using them may be exported and imported using the [Engine Sync feature](#). Classifiers are included when the **Export Settings** action is performed from the **Environments** tab.

ASDD limitations

The primary limitation of the ASDD Profiler is that it is not yet supported for all connectors. The UI will report an error if the user attempts to save a job using an ASDD profile set with an unsupported connector.

Currently, the following conditions must be met to use the ASDD Profiler:

- File profiling is not supported.
- Database connectors using Kerberos authentication are not supported.
- Salesforce SSO authentication is not supported when using an extended connector.
- Discovery may fail or produce lower-quality results for some extended connectors due to known issues:
 - Database variants that add a suffix or prefix to column names will see reduced accuracy for path classifiers. This affects Salesforce as it appends "__c" to the names of user-created columns.

- The SQL syntax used for column truncation is not compatible with all database variants. This is known to cause failures in discovery jobs for the Informix database.

Securing sensitive data

This section contains the following topics:

- [Algorithms](#)
- [Builtin Driver Supports](#)
- [Creating masking jobs](#)
- [Managing jobs](#)
- [Monitoring masking job](#)
- [Masking job wizard](#)
- [Running stopping jobs](#)

Algorithms

Introduction to Masking Algorithms

This article provides a brief outline of the different algorithm options that are available, along with other general algorithm information. More specific algorithm details can be explored in the [Out Of The Box Algorithm Instances](#) or [Algorithm Frameworks](#) sections.

An algorithm plugin can be configured through the graphical user interface by entering the plugin's required configuration in JSON format. For more information, visit the [General UI for Extended Algorithms](#) article.

Algorithm options

Out of the box algorithm instances

Out of the box algorithm instances are pre-configured ready to use algorithms. The out of the box algorithms with related frameworks can be customized using the corresponding extensible frameworks. For more information on algorithm instance extensibility, see [Extensible algorithms](#).

Algorithm Instances	Extensible?	Related Framework
AccNoLookup	X	Secure Lookup
AddrLookup	X	Secure Lookup
AddrLine2Lookup	X	Secure Lookup
BusinessLegalEntityLookup	X	Secure Lookup
dlpx-core:CM Alpha-Numeric	X	Character Mapping
dlpx-core:CM Digits	X	Character Mapping
dlpx-core:CM Numeric	X	
CommentLookup	X	Secure Lookup
Credit Card	X	Payment Card
Date Shift Discrete	X	
Date Shift Fixed	X	Date Shift

Algorithm Instances	Extensible?	Related Framework
Date Shift Variable	X	
DrivingLicenseNoLookup	X	Secure Lookup
DummyHospitalNameLookup	X	Secure Lookup
EmailLookup	X	Secure Lookup
dlpx-core:Email SL	X	Email
dlpx-core:Email Unique	X	Email
dlpx-core:FirstName	X	Name
FirstNameLookup	X	Secure Lookup
dlpx-core:FullName	X	Full Name
FullNMLookup	X	Secure Lookup
LastCommaFirstLookup	X	Secure Lookup
dlpx-core:LastName	X	Name
LastNameLookup	X	Secure Lookup
NullValueLookup	X	
dlpx-core:Phone Unique	X	
dlpx-core:Phone US	X	
RandomValueLookup	X	Secure Lookup
RepeatFirstDigit	X	
SchoolNameLookup	X	Secure Lookup

Algorithm Instances	Extensible?	Related Framework
SecureShuffle	X	
USCitiesLookup	X	Secure Lookup
USstatecodesLookup	X	Secure Lookup
USstatesLookup	X	Secure Lookup
WebURLsLookup	X	Secure Lookup

Algorithm frameworks

Algorithm frameworks allow for creation of algorithm instances with a custom configuration. For more information on algorithm framework extensibility, see [Extensible Algorithms](#). More information on multi-column algorithms can be found at [Using Multi-Column Algorithms](#).

Algorithm Framework	Extensible?	Multi-Column?	Out of the Box Instances
Binary Lookup	X		
Character Mapping	X		dlpx-core:CM Alpha-Numeric dlpx-core:CM Digits
Data Cleansing	X		
Date Replacement	X		
Date Shift	X		Date Shift Fixed
Dependent Date Shift	X	X	
Email	X		dlpx-core:Email Unique dlpx-core:Email SL
Free Text Redaction	X		

Algorithm Framework	Extensible?	Multi-Column?	Out of the Box Instances
Full Name	X		dlpx-core:FullName
Mapping	X		
Min Max	X		
Name	X		dlpx-core:FirstName dlpx-core:LastName
Numeric Expression	X		
Payment Card	X		Credit Card
Regex Decompose	X		
Secure Lookup	X		See Out Of The Box Algorithm Instances > Secure Lookup for all Secure Lookup algorithm instances
Segment Mapping	X		
Tokenization	X		

Configuring your own algorithms

Algorithm settings

The **Algorithm** tab displays algorithm Names along with Type and Description. This is where you add (create) new algorithms. The default algorithms and any algorithms you have defined appear on this tab.

The screenshot shows the 'DELPHIX MASKING' interface. At the top, there's a navigation bar with 'Settings' highlighted. Below it, the breadcrumb 'Home > Settings > Algorithm' is visible. The main heading is 'Settings' with a link to 'Extension support policy'. A 'Nonconforming Data behavior' dropdown is set to 'Mark job as Succeeded', with a 'Learn More' link. A blue 'Add Algorithm' button is in the top right. A table lists various algorithms with columns for Name, Framework, Provider, and Edit/View/Delete.

Algorithms	Name	Framework	Provider	Edit/View/Delete
Custom Algorithms (legacy)	ACCOUNT SL	SL	Built-in	
Domains	ACCOUNT_TK	TA	Built-in	
Profiler	ADDRESS LINE 2 SL	SL	Built-in	
Roles	ADDRESS LINE SL	SL	Built-in	
File Formats	BUSINESS LEGAL ENTITY SL	SL	Built-in	
JDBC Drivers	COMMENT SL	SL	Built-in	

At the top of the page, **Nonconforming Data behavior** is displayed to specify how all algorithms should behave if they encounter data values in an unexpected format. **Mark job as Failed** instructs algorithms to throw an exception that will result in the job failing. **Mark job as Succeeded** instructs algorithms to ignore the non-conformant data and not throw an exception. Note that **Mark job as Succeeded** will result in the non-conformant data not being masked should the job succeed, but the **Monitor** page will display a warning that can be used to report the non-conformant data events.

Creating new algorithms

If none of the default algorithms meet your needs, you might want to create a new algorithm. An algorithm that you create is called a "user-defined algorithm".

Algorithm Frameworks give you the ability to quickly and easily define the algorithms you want, directly on the Settings page. After you create an algorithm, your algorithm will be available to all users.

To add an algorithm:

1. In the upper right-hand corner of the **Algorithm** settings tab, click **Add Algorithm**.

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping (legacy)
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Extended

Create Secure Lookup Algorithm [Learn More](#)

Algorithm Name

Description

Output (Masked) Case

Preserve Lookup File Case

Hash Method

SHA256

Case Sensitive Lookup

Lookup File for Name - Upload a File or Specify a URI

Select...

Cancel

Save

2. Select an algorithm type.
3. Complete the form to the right to name and describe your new algorithm.
4. Click **Save**.

Editing algorithms

Administrators, as well as users with EDIT Algorithm permission assigned in their Role, may edit any user-defined algorithm on the system.

The following algorithm instances cannot be modified:

- Instances that ship with and are defined by the system
- Instances defined by algorithm plugins

Multi-column algorithms

Overview

Multi-column algorithms are a special kind of algorithm that allow a single algorithm assignment to be made spanning multiple columns or fields in inventory. This allows coordinated masking of multiple fields - for example, masking two date-time values while preserving the interval between them.

The [Dependent Date Shift](#) algorithm is an example of a multi-column algorithm.

Usage

Each multi-column algorithm defines a set of **Logical Fields**; these logical fields are assigned to the actual fields or columns in inventory, defining how each value will be treated by the algorithm. A particular logical field may be *read-only*, indicating that it is considered as input but not masked by the multi-column algorithm, and/or *optional*, meaning the logical field is not required in order for the masking assignment to be complete. Furthermore, the **Algorithm Group** number allows a multi-column algorithm to be assigned multiple times in the same table or file-format, with the group number indicating which set(s) of logical fields should be processed together as a single assignment.

 Incomplete multi-column masking assignments in the inventory may not be detected until such time as a masking job is executed using that inventory. It is important to review each multi-column assignment carefully to ensure that for each *Algorithm Group*, each non-optional *Logical Field* is assigned to a column or field in the table or file-format.

Limitations

Multi-column algorithms may only be applied in inventories for data connectors where entire rows or records are processed as a unit.

Specific limitations:

- Multi-column algorithms are not supported for **XML** file masking.
- Multi-column algorithm assignments must be contained with a single Record Type for delimited and fixed-width files.
- Multi-column algorithm assignments must not cross redefines in VSAM copybooks.
- Multi-column algorithms may not be called by other algorithms through the algorithm chaining feature.

Algorithm frameworks overview

Choosing an algorithm framework

See the Algorithm Frameworks section for a detailed description of each Algorithm Framework. The algorithm framework you choose will depend on the format of the data and your internal data security guidelines.

Choosing between character and segment mapping frameworks

The Character Mapping algorithm is intended to replace Segment Mapping for many use cases. That said, it does not replicate every feature of that algorithm, so the specific masking application will determine which one is appropriate.

Reasons to choose Character Mapping over Segment Mapping:

- Character Mapping can mask all characters in the first Unicode plane. Segment Mapping can only mask "[a-zA-Z]" + "[0-9]"
- Character Mapping automatically preserves all non-masked characters. Segment Mapping requires configuration of preserve characters. Character Mapping is much easier to use when the data is potentially "dirty" or not consistently formatted.
- Character Mapping can process preserve ranges in reverse, allowing the last positions of an input to be preserved when inputs have different lengths. Segment Mapping preserve ranges are always processed from the beginning of input.

- Character Mapping uses a more complex masking computation, so that every maskable position influences every other position in the masked value. Segment Mapping pre-computes the permutations for each segment independently.

Reasons to choose Segment Mapping over Character Mapping:

- Segment mapping can mask different parts of the input, determined by position, differently. Character Mapping always masks the same groups of characters regardless of position.
- Segment mapping can map inputs to different outputs at a position, like { A, B, C, D } -> { W, X, Y, Z } by specifying different *Input* and *Mask* values. This is not possible with Character Mapping.
- Segment mapping supports numeric segments, with up to 6-digit segments masked to a specific range. Character Mapping doesn't allow this kind of range limiting.

Out of the box algorithm instances

This section contains the following topics:

- [dlpx-coreCM Alpha-Numeric](#)
- [dlpx-coreCM Digits](#)
- [dlpx-coreCM Numeric](#)
- [Credit Card](#)
- [Date Shift Discrete](#)
- [Date Shift Fixed](#)
- [Date Shift Variable](#)
- [dlpx-coreEmail SL](#)
- [dlpx-coreEmail Unique](#)
- [dlpx-coreFirstName](#)
- [dlpx-coreFullName](#)
- [dlpx-coreLastName](#)
- [NullValueLookup](#)
- [dlpx-corePhone Unique](#)
- [dlpx-corePhone US](#)
- [RepeatFirstDigit](#)
- [Secure Lookup \(Out of the box algorithm instances\)](#)
- [dlpx-coreIBAN](#)
- [SecureShuffle](#)

dlpx-coreCM Alpha-Numeric

Based > **Extensible Algorithm Framework**

The CM Alpha-Numeric algorithm is an instance of the [Character Mapping Algorithm Framework](#).

[-] CM Alpha-Numeric should only be used on non-numeric data types.

This algorithm masks all ASCII digit, lowercase, and uppercase characters, as well as some extended latin and cyrillic characters. Refer to the framework description for details of how masking is performed.

At least one character in the input must be masked, or Non-Conformant data handling will be triggered.

[-] The character mapping algorithm can be used for tokenization and reidentification jobs.

For example:

- "6379315274824970" → "0345698341375224"
- "ABCxyz123" → "HANwhp391"
- "Sí" → "Cž"
- "999-12-3456." → "668-23-1138."
- "2000:a86f::1" → "3893:u55x::0"

[-] This algorithm may generate non-conformant data events.

dlpx-coreCM Digits

Based > Extensible Algorithm Framework

The CM Digits algorithm is an instance of the [Character Mapping Algorithm Framework](#).

This algorithm masks all ASCII digits. Refer to the framework description for details of how masking is performed. Be aware that this algorithm can produce value collisions when applied to Numeric data types. This is because leading zeros are not significant in numeric types, so while "7" → "8" and "304" → "008" may be different string results, when inserted into a numeric field, they represent the same value. If this behavior is undesirable, consider using the [CM Numeric](#) algorithm.

At least one character in the input must be masked, or Non-Conformant data handling will be triggered.

For example:

- "6379315274824970" → "8345698341375224"
- "99" → "05"
- "ABCxyz123" → "ABCxyz391"
- "0" → "6"



This algorithm may generate non-conformant data events.

dlpx-coreCM Numeric

Based > Extensible Algorithm Framework

The CM Numeric algorithm is an algorithm based on logic in the [Character Mapping Algorithm Framework](#).

 CM Numeric algorithm should only be used for numeric data types.

The framework this algorithm is based on is not configurable and cannot be reused to create additional instances.

This algorithm masks all ASCII digit without the possibility of the first digit masking to "0". Leading and trailing zeros are preserved. The value "0" always masks to "0". Unlike the "CM digits" instance, the number of significant digits is always preserved for all numeric inputs.

Refer to the framework description for details of how masking is performed.

At least one character in the input must be masked, or Non-Conformant data handling will be triggered.

 This algorithm can only be used for integer data in tokenization and reidentification jobs. Masking numbers with decimal points is not reversible.

For example:

- "6379315274824970" → "5210366768740261"
- "99" → "75"
- "000051.1230" → "000072.9040"
- "ABCxyz123" → "ABCxyz391"
- "0" → "0"

 This algorithm may generate non-conformant data events.

Credit Card

Based > Extensible Algorithm Framework

The Credit Card algorithm is an instance of the [Payment Card Algorithm Framework](#). The algorithm requires input values to have at least 8 digits in the character group [0-9]. If an input value has less than this, the algorithm will return an error. It preserves the first 6 digits of the input and requires at least one position to be masked for masking to be considered successful. The algorithm masks all subsequent digits by replacing them with a random value. All input characters that are not in the character group [0-9] are preserved. The algorithm maintains Luhn check validity through masking so input values with a valid Luhn check will mask to a value with a valid Luhn check. The out-of-the-box instance of this algorithm is called **CreditCard**.

For example:

- "6379315274824970" → "6379318341375224"
- "6379.3152.7482.4970" → "6379.3183.4137.5224"
- "abc5473defg04828hijkl0656253" → "abc5473defg04971hijkl6490341"



This algorithm may generate non-conformant data events.

Date Shift Discrete

The Date Shift Discrete algorithm masks all dates with the same year-month combination to the same day. A different day is returned for each year-month combination. As an example, any inputs with a year-month combination of February 2020 may return a day value of 23 while any inputs with a year-month combination of January 2020 may return a day value of 5. All values of the input other than the day value are preserved. This algorithm is deterministic based on an algorithm key. The out-of-the-box instance of this algorithm is called **DateShiftDiscrete**.

For example:

- "1989-11-19 00:00:00" → "1989-11-30 00:00:00"
- "1989-12-19 04:15:00" → "1989-12-24 04:15:00"
- "2012-11-19 17:00:55" → "2012-11-08 17:00:55"
- "2012-11-09 00:23:59" → "2012-11-08 00:23:59"



This algorithm may generate non-conformant data events.

Date Shift Fixed

Based > Extensible Algorithm Framework

The Date Shift Fixed algorithm is an instance of the [Date Shift Algorithm Framework](#) masking the input to 5 days in the future with roll enabled so only the day of the month will change, all other units will remain the same. Dates at the end of the month will roll back to the beginning of the same month in the same year. The out-of-the-box instance of this algorithm is called **DateShiftFixed**.

For example:

- "2001-02-05 12:30:00" → "2001-02-10 12:30:00"
- "2001-02-27 15:45:00" → "2001-02-04 15:45:00"
- "2001-12-28 00:00:00" → "2001-12-02 00:00:00"



This algorithm may generate non-conformant data events.

Date Shift Variable

The Date Shift Variable algorithm returns a random date within the same month-year as the input date. Dates will not mask to the original input date. This algorithm may produce collisions. The out-of-the-box instance of this algorithm is called **DateShiftVariable**.

For example:

- "2019-02-05 10:00:00" → "2019-02-13 10:00:00"
- "2019-02-12 15:30:00" → "2019-02-13 15:30:00"
- "2019-02-27 00:45:30" → "2019-02-17 00:45:30"
- "2020-02-27 00:00:00" → "2020-02-22 00:00:00"



This algorithm may generate non-conformant data events.

dlpx-coreEmail SL

Based > Extensible Algorithm Framework

The Email SL algorithm is an instance of the [Email Algorithm Framework](#). This algorithm splits the input on the '@' symbol. [Handling of malformed inputs](#) is detailed on the [Email Algorithm Framework page](#). This algorithm does not generate any non-conformant data events. The algorithm will split the input into two parts: **name** and **domain**. Name is the portion before the '@' symbol and domain is the portion after the '@' symbol.

A secure lookup is applied to the name portion of the input. The provided secure lookup file contains 20,000 unique lookup values in various formats. The following formats are used in the default lookup file:

- FirstName.LastName
- FirstName_LastName
- FirstInitial.LastName
- FirstNameLastName
- FirstNameLastInitialNumber

The domain portion is replaced by the fixed value "example.com". This value is a reserved domain with a valid DNS entry.

This algorithm is deterministic based on an algorithm key. It is possible that there may be collisions where two different values mask to the same value due to the nature of secure lookup. The out-of-the-box instance of this algorithm is called **dlpx-core:Email SL**.

For example:

- "bob@gmail.com" → "E.Duboise@example.com"
- "bob@hotmail.com" → "E.Duboise@example.com"
- "alex@gmail.com" → "OrvinA436@example.com"
- "joe_123@yahoo.com" → "Amil.Steidinger@example.com"

dlpx-coreEmail Unique

Based > Extensible Algorithm Framework

The Email Unique algorithm is an instance of the [Email Algorithm Framework](#). This algorithm splits the input on the '@' symbol. [Handling of malformed inputs](#) is detailed on the [Email Algorithm Framework](#) page. This algorithm does not generate any non-conformant data events. The algorithm will split the input into two parts: **name** and **domain**. Name is the portion before the '@' symbol and domain is the portion after the '@' symbol.

The name portion is masked by performing a SHA-256 hash of the entire input (including the domain). This means that inputs with the same name portion but different domain portions will mask to different values. The hashed value is then encoded using Base32 encoding. The result of these transformations is the masked name portion.

 This instance may produce masked name portions with lengths up to 52 characters.

The domain portion is replaced by the fixed value "example.com". This value is a reserved domain with a valid DNS entry.

This algorithm is deterministic based on an algorithm key. This algorithm provides unique masked values for each input. The out-of-the-box instance of this algorithm is called **dlpx-core:Email Unique**.

For example:

- "bob@gmail.com" → "XF35TNMKPPTMQF4CX5264ZRXOMJJL2DQVE3KTZNIJ2NS6EUH7GLA@example.com"
- "bob@hotmail.com" → "M2U3LCC24MP5XDQ7DH4RSDW6QXCWRTSJVF22C7IKBXDQ3LBM7NQ@example.com"
- "alex@gmail.com" → "CQKOVBP3VT42XHLBBUHEWIAJ26X3NROEBZHM7SC7B4NFSZSTBIQ@example.com"
- "joe_123@yahoo.com" → "JTJNSLW4TK4TWQ7VKG2KMRMMH4M3FRIXUXFR7TIEL6VJR3G6AU2Q@example.com"

dlpx-coreFirstName

Based > Extensible Algorithm Framework

The First Name algorithm is an instance of the [Name Algorithm Framework](#). The algorithm requires String type input values.

The expected format for the valid input contains at least one word, which consist of at least one non-whitespace character. If the input value does not match the expected format, the value will not be masked. I.e. if input contains null or empty string or white spaces only, then the algorithm returns unmasked input value.

No non-conformant data errors are thrown by that algorithm.

Single character is considered an abbreviation, i.e. it will be masked to a single character. Whether it is followed by the dot (.) or not. Words separated by the hyphen (-) are considered as a single word (even if divided from hyphen by spaces).

The default First Name instance is configured without particle files, so every input word is considered as a valid part of the name. The maximum number of names masked and returned is 2; additional names are dropped. Leading and trailing white spaces are not preserved.

For example:

Input	Masked Output
null	null
"" (empty string)	""
" " (white spaces only)	" "
& (single non alphanumeric character)	R
&?	Michael
M	S
M.	S.
Ann- Marie	Boris
Ann-Marie M.	Boris S.
Ann-Marie Jane Jill	Boris Shelby
von (particle not configured)	tim
von Jane	tim Shelby

dlpx-coreFullName

Based > Extensible Algorithm Framework

The Full Name algorithm is an instance of the [Full Name Algorithm Framework](#). The algorithm requires String type input values.

If input value is non-conformant (for example: null or white spaces), it is not masked. Words containing any character(s) are considered as a valid input and masked. Words separated by hyphen (-) are considered as a single word (even if divided from hyphen by spaces). No non-conformant data errors are thrown by that algorithm.

The default Full Name algorithm instance uses all default parameters, and chains "dlpx-core:FirstName" algorithm instance for first names masking, and "dlpx-core:LastName" for last name masking.

Below are few examples of the Full Name default algorithm instance masking:

Input	Masked Output
Manuel Maria Saxe-Coburgo-Gotha	Nimisha Kum Mcneish
Manuel - Boris Maria Saxe-Coburgo-Gotha	Simeon Kum Mcneish
Manuel Maria Saxe -Coburgo - Gotha	Nimisha Kum Mcneish
Manuel Maria de Saxe-Coburgo-Gotha	Nimisha Kum Mcneish
Manuel Maria de Saxe-Coburgo-Gotha (*)	Nimisha Kum Casteleyn
Manuel Maria - de ? Saxe-Coburgo-Gotha : #	Nimisha Muharrem E
Manuel Maria Saxe-Coburgo-Gotha (*)	Nimisha Kum Casteleyn
Mr. Manuel Maria de Saxe-Coburgo-Gotha #	Levar Nimisha E (dlpx-core:FirstName does not have any configured particles)
Saxe-Coburgo-Gotha, Manuel Maria	Mcneish, Nimisha Kum
saxe-coburgo-gotha, Manuel Maria	mcneish, Nimisha Kum
SAXE-COBURGO-GOTHA, MANUEL Maria	MCNEISH, NIMISHA Kum
Saxe-Coburgo-Gotha: M. M	Claudia T. S
M. G. Maria Saxe-Coburgo-Gotha	T. E. Mcneish

Input	Masked Output
M M Saxe-Coburgo-Gotha	T T Mcneish
M M. Saxe-Coburgo-Gotha	T T. Mcneish
M M. S	T T. G
M M. S.	T T. G.
m m. s.	t t. g.
Max	Grassi
Max	Grassi

dlpx-coreLastName

Based > Extensible Algorithm Framework

The Last Name algorithm is an instance of the [Name Algorithm Framework](#). The algorithm requires String type input values.

The expected format for the valid input contains at least one word, which consist of at least one non-whitespace character. If the input value does not match the expected format, the value will not be masked. I.e. if input contains null or empty string or white spaces only then the algorithm returns unmasked input value.

No non-conformant data errors are thrown by that algorithm.

Single word input is not checked for configured particles. Single character is considered abbreviation, i.e. it will be masked to a single character, whether it is followed by the dot (.) or not. Words separated by hyphen (-) are considered as a single word (even if divided from hyphen by spaces).

The default Last Name instance is configured with a particleToRemove file. After configured particles are removed from the input, a single word is masked and returned; additional words are dropped. Leading and trailing white spaces are not preserved.

For example:

Input	Masked Output
null	null
"" (empty string)	""
" " (white spaces only)	" "
? (single non alphanumeric character)	K
?>	Michael
M	S
M.	S.
Ann- Marie	Boris
von (particle)	wilke
FROUT	Smith
von FROUT	Smith
von FROUT Weissman	Smith

NullValueLookup

Based > Extensible Algorithm Framework

This algorithm replaces the input with a null or empty value, depending on the context.

 The algorithm's name is chosen for backward compatibility only. It does not perform any kind of lookup and is not related to the Secure Lookup framework.

For example:

- "6379315274824970" → null
- "ABCxyz123" → null
- "Sí" → null
- "999-12-3456." → null
- "2000:a86f::1" → null

dlpx-corePhone Unique

Based > Extensible Algorithm Framework

The Phone Unique algorithm masks the last 7 digits in the character group [0-9] with the hash value of the digits. All characters outside of this character group remain unmasked and are preserved in the masked value.

The maximum acceptable input length is 30 symbols, longer inputs will trigger non-conformant data handling. The input must contain at least one character in the character group [0-9], or non-conformant data handling will be triggered.

For example:

- "12-765" → "29-540"
- "(123)456-7890" → "(123)012-3901"
- "1(800) FLOWERS" → "2(746) FLOWERS"
- "+1-650-513-0514" → "+1-650-409-9747"
- "(512) 333-1234 ext 123" → "(512) 333-2905 ext 908"
- "CALL-ME-FLOWERS" → "CALL-ME-FLOWERS" (and generates a non-conformant data event)



This algorithm may generate non-conformant data events.

dlpx-corePhone US

Based > Extensible Algorithm Framework

The Phone US algorithm masks the last 4 digits in the character group [0-9] with the hash value of the digits and the 3 preceding digits are replaced with the value '555'. All characters outside of this character group remain unmasked and are preserved in the masked value.

The maximum acceptable input length is 30 symbols, longer inputs will trigger non-conformant data handling. The input must contain at least one character in the character group [0-9], or non-conformant data handling will be triggered.

For example:

- "12-765" → "58-504"
- "(123)456-7890" → "(123)555-3085"
- "1(800) FLOWERS" → "2(746) FLOWERS"
- "+1-650-513-0514" → "+1-650-555-9202"
- "(512) 333-1234 ext 123" → "(512) 333-5550 ext 497"
- "CALL-ME-FLOWERS" → "CALL-ME-FLOWERS" (and generates a non-conformant data event)



This algorithm may generate non-conformant data events.

RepeatFirstDigit

Based > Extensible Algorithm Framework

This algorithm masks the "+4" component of a zip code by repeating its first digit four times, unless the first digit is zero, in which case '1' is repeated four times.

The input must contain a 4-character string "DDDD" (where each 'D' is a numeric digit). The following formats are valid inputs:

- 4-character string "DDDD" where 'D' is a digit
- 9-character string "ccccDDDD" where 'D' is a digit, and 'c' can be any character
- 10-character string "ccccDDDD" where 'D' is a digit, and 'c' can be any character but must contain at least one hyphen or period
- 14-character string "ccccDDDDcccc" where 'D' is a digit, and 'c' can be any character

 14-character "ccccDDDDcccc" inputs will be truncated to "ccccDDDD"

For example:

- "6912" → "6666"
- "0123" → "1111"
- "941173564" → "941173333"
- "43556-9703" → "43556-9999"
- "009078377 SJPR" → "009078888"

 This algorithm may generate non-conformant data events.

Secure Lookup (Out of the box algorithm instances)

This section covers the following topics:

- [AccNoLookup](#)
- [AddrLookup](#)
- [AddrLine2Lookup](#)
- [BusinessLegalEntityLookup](#)
- [CommentLookup](#)
- [DrivingLicenseNoLookup](#)
- [DummyHospitalNameLookup](#)
- [EmailLookup](#)
- [FirstNameLookup](#)
- [FullNMLookup](#)
- [LastCommaFirstLookup](#)
- [LastNameLookup](#)
- [RandomValueLookup](#)
- [SchoolNameLookup](#)
- [USCitiesLookup](#)
- [USCountiesLookup](#)
- [USStatecodesLookup](#)
- [USStatesLookup](#)
- [WebURLsLookup](#)

AccNoLookup

Based > Extensible Algorithm Framework

The AccNoLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are 5 digit account numbers.

For example:

- "6379315274824970" → "64893"
- "ABCxyz123" → "72345"
- "ID3938491" → "72433"
- "999-12-3456" → "25326"
- "2000:a86f::1" → "86432"

AddrLookup

| **Based >Extensible Algorithm Framework**

The AddrLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are line one address values.

For example:

- "49 Main St" → "55 BLUE DR"
- "1947 Highway 5" → "92 GREEN ST"
- "9 County Route 52.5" → "1049 ORANGE CIRCLE"

AddrLine2Lookup

Based > Extensible Algorithm Framework

The AddrLine2Lookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are line two address values such as apartment number.

For example:

- "#483" → "UNIT 29"
- "APT 3D" → "P.O. BOX 934"
- "unit 13B" → "APARTMENT 1"

BusinessLegalEntityLookup

| **Based > Extensible Algorithm Framework**

The BusinessLegalEntityLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are legal business names.

For example:

- "XYZ Corp." → "Boeing"
- "Alpha LLC" → "3M"
- "ABC Inc." → "Campbell Soup"

CommentLookup

Based > Extensible Algorithm Framework

The CommentLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm. All non-empty and non-null inputs to this algorithm will mask to the same value.

The lookup value for this algorithm is a generic comment value.

For example:

- "6379315274824970" → "This data has been masked in all non-production environments as per Enterprise Information Security Policy(2013)."
- "ABCxyz123" → "This data has been masked in all non-production environments as per Enterprise Information Security Policy(2013)."
- "Sí" → "This data has been masked in all non-production environments as per Enterprise Information Security Policy(2013)."
- "999-12-3456." → "This data has been masked in all non-production environments as per Enterprise Information Security Policy(2013)."
- "2000:a86f::1" → "This data has been masked in all non-production environments as per Enterprise Information Security Policy(2013)."

DrivingLicenseNoLookup

| **Based > Extensible Algorithm Framework**

The DrivingLicenseNoLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are 9-digit driver's license IDs.

For example:

- "6379315274824970" → "865345234"
- "ABCxyz123" → "952731585"
- "US949382" → "164927562"

DummyHospitalNameLookup

| **Based >Extensible Algorithm Framework**

The DummyHospitalNameLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are non-real hospital names.

For example:

- "Hospital 1" → "Community Hospital"
- "New York General Hospital" → "St. Patrick's Medical Center"
- "California Health Institute" → "Gotham City Mental Hospital"
- "Children's Hospital of Philadelphia" → "Hogwarts Medical Clinic"

EmailLookup

Based > Extensible Algorithm Framework

The EmailLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

formation.



A new email framework and two new email algorithm instances were introduced in version 6.0.9.0 and are the preferred methods for masking email values. See [Email](#), [Email SL](#), and [Email Unique](#) for more in

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are non-resolving email addresses.

For example:

- "bob@gmail.com" → "Andy.Samberg@nytimes.edu"
- "Albert_Einstein@nasa.gov" → "John.Smith@aol.gov"
- "abc123@delphix.com" → "Fred.James@yahoo.net"

FirstNameLookup

Based > Extensible Algorithm Framework

The FirstNameLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

 A new name framework and a new first name algorithm instance was introduced in version 6.0.8.0 and are the preferred methods for masking name values. See [Name](#) and [FirstName](#) for more information.

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are first names.

For example:

- "lucas" → "Jacob"
- "Gabby Elizabeth" → "Dennis"
- "John Jacob Jingleheimer Schmidt" → "Ray"

FullNMLookup

Based > Extensible Algorithm Framework

The FullNMLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

 A new full name framework and a new full name algorithm instance was introduced in version 6.0.8.0 and are the preferred methods for masking full name values. See [FullName\(framework\)](#) and [FullName\(instance\)](#) for more information.

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are full names (first & last name).

For example:

- "Harry Potter" → "John Wick"
- "joe" → "Carol Reed"
- "Robert Downey Jr." → "Aaron Burr"
- "Queen Elizabeth II" → "Ferris Bueller"

LastCommaFirstLookup

Based > **Extensible Algorithm Framework**

The LastCommaFirstLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

 A new full name framework was introduced in version 6.0.8.0 and is the preferred methods for masking full name values. See [Full Name](#) for more information.

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are full names in the format Last Name, First Name.

For example:

- "Lincoln, Abe" → "Campbell, Allison"
- "George Washington" → "Douglas, Alfred"
- "teddy" → "Smith, Jack"

LastNameLookup

Based > Extensible Algorithm Framework

The LastNameLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

 A new name framework and a new last name algorithm instance was introduced in version 6.0.8.0 and are the preferred methods for masking name values. See [Name](#) and [LastName](#) for more information.

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are last names.

For example:

- "Smith" → "Blair"
- "santa-cruz" → "Carney"
- "von Trapp" → "Washington"

RandomValueLookup

Based > Extensible Algorithm Framework

The RandomValueLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are 50-character random value strings.

For example:

- "6379315274824970" → "ufGcYiFgzC6RBmcBPC7Mvb0oOqEhjEVDUJZLHo6OYNWoi5PZKC"
- "ABCxyz123" → "Wk4iq8Y6Ngz8j84AhueDmHQo6uQgMjmnMLuGFxuPEPmDBLzzNf"
- "Sí" → "8pz8lnVeQqNCe3jnQREPH2bfbQHkNRir6CHliwq1fMTY3sKFIY"
- "999-12-3456." → "37cq1Lve2rOi3kwrNjDE2p1CNPSeAUtnZYNRfUcDuGnix6DE9"
- "2000:a86f::1" → "UUWWeetRXJXMR6puAk8414nrHwMn5nanrOxoWw7DesbHHZPLs1"

SchoolNameLookup

| **Based > Extensible Algorithm Framework**

The SchoolNameLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are schools and colleges in the format School Name, State, USA.

For example:

- "Columbia University" → "Smith College, Massachusetts, USA"
- "clown college" → "Ithaca College, New York, USA"
- "The Culinary Institute" → "Stanford University, California, USA"
- "UCLA" → "Rensselaer Polytechnic Institute, New York, USA"

USCitiesLookup

| **Based >Extensible Algorithm Framework**

The USCitiesLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are United States cities.

For example:

- "Los Angeles" → "Chicago"
- "New England" → "Oklahoma City"
- "cape town, south africa" → "Houston"
- "Princeton, New Jersey" → "Redwood City"

USCountiesLookup

| **Based > Extensible Algorithm Framework**

The USCountiesLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are United States counties.

For example:

- "Schuyler County" → "Rock Island County"
- "orange co." → "Price County"
- "Yellowstone County, Montana" → "McKinley County"

USstatecodesLookup

| **Based >Extensible Algorithm Framework**

The USstatecodesLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are 2-letter United States state and territory codes.

For example:

- "AL" → "CA"
- "Maine" → "UT"
- "west virginia" → "SD"
- "Italy" → "PR"

USstatesLookup

| **Based >Extensible Algorithm Framework**

The USstatesLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are United States state and territory names.

For example:

- "Hawaii" → "Iowa"
- "KS" → "District of Columbia"
- "california" → "Northern Marianas Islands"

WebURLsLookup

| **Based > Extensible Algorithm Framework**

The WebURLsLookup algorithm is an instance of the [Secure Lookup Algorithm Framework](#).

This algorithm performs a lookup on the input value and returns a value from the provided lookup file. It is possible for this algorithm to produce the same output value for different input values. Inputs to this algorithm are case-sensitive so two inputs with the same value in different cases may mask to different values. Leading and trailing whitespaces are preserved by this algorithm.

The lookup values for this algorithm are website addresses.

For example:

- "www.google.com" → "http://www.blogspot.com"
- "delphix.com" → "http://www.gaurdian.co.uk"
- "https://en.wikipedia.org/wiki/Syslog#References" → "http://www.newegg.com"

dlpx-coreIBAN

[Extensible Algorithm Framework](#)

The IBAN algorithm allows you to mask [International Bank Account Numbers](#). It takes IBAN as an input and replaces each numeric character with another numeric character, and the same for alphabet characters. The first two characters of a valid IBAN are its country code. The country code is preserved, it is not masked. The second two characters of a valid IBAN are a checksum, generated from the other characters using the MOD97 method. This is regenerated by the algorithm after masking to ensure checksum validity. The replacement characters are sourced from a hash of the original input IBAN. This hash remains stable for an instance of the algorithm, so the output will always be the same for a given input. A small set of delimiters are preserved if present in the input IBAN: space, period, and hyphen. Delimiters will not be preserved in the first 4 characters of an input, as these are reserved for country code and checksum.

The algorithm attempts to mask all inputs even if they are not valid IBAN and will still replace numeric characters with numeric, and alphabet with alphabet. Unsupported delimiters or punctuation will be replaced with an alphanumeric character. If possible, a new checksum is calculated even if the input did not have a valid checksum. The algorithm cannot proceed with masking invalid data that has less than 4 characters. If this input is received, it will be padded to 4 characters and masking will reattempt.

Creating an Instance / Configuration Options

The default instance of the IBAN algorithm cannot be configured. Newly created instances of the IBAN algorithm have two options that can be configured. As with other algorithms, enter a name and an optional description. You can then choose to enable input validation and select how many characters to mask.

The number of characters to mask maximum is 64, but the longest country supported IBAN is currently Russia with 33 characters. Numbers higher than the length of an IBAN input will mask all characters possible. The minimum number is 6. Lower numbers will speed up the performance of the algorithm, but be conscious of reduction in data obfuscation as the number of changed characters goes down.

There is an option for IBAN input validation. This can be useful to identify corrupted or otherwise invalid input data. It does come at a moderate performance cost. The validation uses the [Apache Commons Validator](#). This tool validates that the country code is supported, the checksum is valid and that the structure of the IBAN is appropriate for the given country. It is updated with new countries as they adopt the IBAN standard. The current version used is 1.7. If invalid IBAN are found, they are treated as non-conformant data. **This will not stop the masking job and the value will not be masked** but a warning will be shown and details recorded in the job log.

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Numeric Expression
- IBAN**
- Extended

Create IBAN Algorithm [Learn More](#)

Algorithm Name

Description

Mask Type
STRING

Validate Input IBAN

Number of Characters to Mask (min: 6, max: 64)
36

Cancel Save

1. In the upper right-hand region of the **Algorithm** tab under **Settings**, click **Add Algorithm**.
2. Select **IBAN**. The "Create IBAN Algorithm" window will appear.
3. Enter an **Algorithm Name**.
Info: Name MUST be unique.
4. Enter a **Description** (optional).
5. Set **Number of Characters to Mask**. This value is the number of characters from the right that will be masked. Values for this field must be in the range [6-64].
6. When finished, click **Save**.

Examples

Valid IBAN inputs and masked outputs:

- NO8330001234567 → NO0631216940542
- GL8964710123456789 → GL3640635860760239

Invalid IBAN inputs and masked outputs:

- ???A???b???C???d???E??? → ??29S1076R7076M0100E0222J0177
- ABC123 → AB6447

Notice that it still respects a character's type and the first two as if they were a country code.

SecureShuffle

This algorithm masks by shuffling the values in a particular field or column to different lines or rows. For example, values for the FIRST_NAME column might be shuffled among a number of database rows within a table. It guarantees that each value is moved to a different line or row, but will not prevent an input from masking to the same output in the case where the values shuffled are not unique.

 Because shuffling data does not redact or modify the individual data values in any way, careful consideration must be given to whether this form of obfuscation is sufficient to meet your security requirements.

The SecureShuffle algorithm may only be used with masking jobs that support batching, and will not be presented as an option in the inventory screen when it is not supported. The maximum number of positions any particular value will be moved within the input is equal to the batch size.

Please refer to the Batch Masking section [here](#) for a full description of the Batch Masking mechanism, as well as details on batch size and which jobs support batching.

This algorithm will report non-conformant data whenever only one value is available to mask, meaning that no shuffling is possible.

Algorithm frameworks

This section covers the following topics:

- [Binary Lookup\(Algorithm frameworks\)](#)
- [Character Mapping \(Algorithm frameworks\)](#)
- [Data Cleansing \(Algorithm frameworks\)](#)
- [Date Replacement \(Algorithm frameworks\)](#)
- [Date Shift \(Algorithm frameworks\)](#)
- [Dependent Date Shift \(Algorithm frameworks\)](#)
- [Email \(Algorithm frameworks\)](#)
- [Free Text Redaction \(Algorithm frameworks\)](#)
- [Full Name \(Algorithm frameworks\)](#)
- [Mapping \(Algorithm frameworks\)](#)
- [Min Max \(Algorithm frameworks\)](#)
- [Name \(Algorithm frameworks\)](#)
- [Numeric Expression \(Algorithm frameworks\)](#)
- [Payment Card \(Algorithm frameworks\)](#)
- [Regex Decompose \(Algorithm frameworks\)](#)
- [Secure Lookup \(Algorithm frameworks\)](#)
- [Segment Mapping \(Algorithm frameworks\)](#)
- [Tokenization \(Algorithm frameworks\)](#)

Binary Lookup(Algorithm frameworks)

Extensible Algorithm Framework

A Binary Lookup algorithm is much like the Secure Lookup algorithm but is used when entire files are stored in a specific column. This algorithm replaces objects that appear in object columns. For example, if a bank has an object column that stores images of checks, you can use a Binary Lookup algorithm to mask those images. The Delphix Engine cannot change data within images themselves, such as the names on X-rays or driver's licenses. However, you can replace all such images with a new, fictional image. This fictional image is provided by the owner of the original data.

Creating a Binary Lookup algorithm via UI

1. At the top right of the **Algorithm** tab, click **Add Algorithm**.
2. Select **Binary Lookup Algorithm**. The Create Binary SL Algorithm pane appears.
3. Enter an **Algorithm Name**.
Info: This MUST be unique.
4. Enter a **Description**.
5. Select a **Binary Lookup File** on your filesystem. A maximum of **50 Lookup Files** can be selected.
6. Click **Save**.

For information on creating Binary Lookup algorithms through the API, see [API Calls for Creating Algorithms - Binary Lookup](#).

Character Mapping (Algorithm frameworks)

Extensible Algorithm Framework

The Character Mapping framework maps text values, defined by a set of character groups, to other text values generated from the same character groups. Mappings are calculated algorithmically, so it is not necessary to provide the set of mapping values. The algorithm preserves any characters not assigned to a group. Any characters from the first Unicode plane can be mapped - this covers most characters used in modern languages. Other (supplementary) characters can only be preserved.

The particular set of permutations used is determined by the algorithm's key, so rekeying the algorithm will cause different outputs to be generated for each input.

The algorithm has the following properties:

- The masked value for each input is consistent unless the algorithm is rekeyed.
- No two text inputs produce the same text output. Collisions *are* possible for some data types, such as Numeric, where multiple text values, such as "001" and "1", are treated as the same value.
- As long as at least one maskable character is present in the input, the masked value will never match the input.
- Each masked position influences the mapping done at every other masked position.

For these reasons, this algorithm is useful for masking columns with uniqueness requirements, such as primary and foreign key columns.

This algorithm was introduced in version 6.0.5.0, and uses the algorithm extensibility framework, allowing it to be called from other algorithms using that framework.

To decide whether Character Mapping or Segment Mapping is the correct option for your use case, see [Choosing Between Character and Segment Mapping Frameworks](#).



The character mapping algorithm can be used for tokenization and reidentification jobs.

Creating a character mapping algorithm via UI

Select Framework

- Secure Lookup
- Character Mapping**
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping (legacy)
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Extended

Create Character Mapping Algorithm

Algorithm Name

Description

Character Groups [Learn More](#)

Case Sensitive

Minimum Masked Positions

Preserve Leading Zeros

Preserve Ranges

Starting Position	Length	Direction	
<input style="width: 80px;" type="text"/>	<input style="width: 80px;" type="text"/>	<input style="width: 80px;" type="text" value="Forward"/>	<input style="margin-left: 10px;" type="button" value="Add"/>

1. In the upper right-hand region of the **Algorithm** tab under **Settings**, click **Add Algorithm**.
2. Select **Character Mapping Algorithm**. The "Create Character Mapping Algorithm" pane appears.
3. Enter an **Algorithm Name**.
Info: This MUST be unique.
4. Enter a **Description**.
5. Define **Character Groups** for each group of characters among which you would like to map. Each group may be defined either by specifying each literal character in the group, such as "0123456789", or using Java Regular Expression style character ranges, such as "[0-9]". The algorithm will freely map characters to other characters within the same group, so by defining groups "[0-9]" and "[A-Z]", numbers would be replaced by other numbers, and letters by other letters, but a number would never be replaced by a letter. Groups should not contain duplicate characters, and each character may belong to only one group. Any character that is not assigned to a group will be preserved (not masked) by the algorithm. The box below the entry area allows selection of character groups defined for other, preexisting Character Mapping algorithms.
6. Check the **Case Sensitive** box to cause the algorithm to treat upper and lower case characters as distinct characters for mapping.
7. Select a value for **Minimum Masked Position**, which sets the minimum number of characters that the algorithm must mask; fewer positions triggers non-conformant data handling. Null, empty, and all-whitespace values never trigger non-conformant data handling.

8. Check the **Preserve Leading Zeros** box to cause the algorithm to preserve any number of '0' characters at the beginning of each input. This is only useful if '0' has been assigned to a character group in step 5.
Warning: Masked results are not guaranteed to be unique if **Preserve Leading Zeros** is used, and the algorithm cannot be used for tokenization/re-identification jobs.
9. If desired, define ranges of the input value to ignore using the **Preserve Ranges** controls. For Character Mapping algorithms, only characters that would otherwise be masked count when determining position for preserve ranges. Each preserve range is defined by:
 - a. **Start Position** - The position at which to start preserving, starting from 0.
 - b. **Length** - The number of characters to preserve.
 - c. **Direction** - The direction, either forward or reverse, determining whether to process from the beginning or end of input for this range.

Examples

As an example, a Character Mapping algorithm could be defined with a single character group, "[0-9]". It might mask as follows:

- "(603) 867-5309" → "(463) 638-0193"
- "999-12-3456" → "453-71-6283"
- "Call Tom at 8:00PM" → "Call Tom at 2:75PM"

Data Cleansing (Algorithm frameworks)

Extensible Algorithm Framework

A data cleansing algorithm is used to standardize varied spellings, misspellings, and abbreviations for the same name. For example, “Ariz,” “Az,” and “Arizona” can all be cleansed to “AZ.” Use this algorithm if the target data needs to be in a standard format prior to masking.

Creating a data cleansing algorithm via UI

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping (legacy)
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Extended

Create Data Cleansing Algorithm [Learn More](#)

Algorithm Name

Description

Case Sensitive Lookup

Trim Whitespace

Lookup File - Upload a File or Specify a URI

Lookup File Delimiter

1. Enter an **Algorithm Name**.

Info:

This MUST be unique.

2. Enter a **Description** (optional).
3. Choose whether to use **Case Sensitive Lookup**. If this box is checked, the data to be cleansed must match the case of the value in the lookup file in order to be replaced.

For example, if the lookup file contains `Arizona=AZ` :

Original	Cleansed	Case Sensitive Lookup
Arizona	AZ	checked or not checked
arizona	AZ	not checked

Original	Cleansed	Case Sensitive Lookup
arizona	arizona	checked

- Choose whether to **Trim Whitespace**. If this box is checked, the leading and trailing whitespace of the data to be cleansed is removed prior to checking if the value is in the lookup file. This allows a single `value=replacement` in the lookup file to cleanse data containing extraneous leading and trailing whitespace.

Info

This must be checked to cleanse fixed-width files and fixed-length database data types such as CHAR and NCHAR.

- Specify a **Lookup File**. You can either click the Select... button to choose a local file or enter the fileReferenced value returned from the fileUpload API endpoint for uploading files to the Masking Engine. The file should contain a newline separated list of {value, replacement} pairs separated by the delimiter.
- Specify a **Lookup File Delimiter** (value and replacement separator) up to 50 characters long. The default delimiter is `=`. You can change this to match the lookup file.
- Click **Save**.

Below is an example of a lookup file. It does not require a header. Make sure there are no spaces or returns at the end of the last line in the file. The following is sample file content:

```
NYC=NY
NY City=NY
New York=NY
Manhattan=NY
```

For information on creating Data Cleansing algorithms through the API, see [API Calls for Creating Algorithms - Data Cleansing](#).

Date Replacement (Algorithm frameworks)

Extensible Algorithm Framework

The Date Replacement framework masks a date value based on specified beginning and end dates. Masked output values are calculated algorithmically using the algorithm's key, so rekeying the algorithm will cause a different output value to be generated for each input. It is possible for an input to be masked to itself.

Creating a Date Replacement Algorithm via UI

1. In the upper right-hand region of the **Algorithm** tab under **Settings**, click **Add Algorithm**.
2. Select **Date**. The "Create Date Algorithm" pane appears.
3. Enter an **Algorithm Name**.
Info: This MUST be unique.
4. Enter a **Description**.
5. Under **Select Algorithm type** choose **Replacement**.
6. Enter **Min Date** and **Max Date**. These define the range from which the algorithm will select output values. The range is inclusive of both values. All units of time less than the specified unit must be set to 0. For example, a configuration with the unit set to **Days** must have the time portion set to 00:00:00.
7. Choose the **Unit** of time form the drop-down: **Days**, **Hours**, **Minutes**, or **Seconds**. This represents the unit of time the range is expressed in. Any unit smaller than the specified unit will be set to 0 in the masked output. For example, with a unit of **Days**, all masked time values will be 00:00:00. For a more detailed explanation, see the [Examples](#) section.
8. When you are finished, click **Save**.

For information on creating Date Replacement algorithms through the API, see [API Calls for Creating Algorithms - Date Replacement](#).

Examples

As an example, a Date Replacement algorithm with a minimum range of "2020-01-01 00:00:00" and a maximum range of "2020-01-05 00:00:00" with the unit set to **Days** will replace the input value with a date in the specified range. Dates may mask as follows:

- "1995-03-05 13:25:00" → "2020-01-02 00:00:00"
- "2021-10-13 01:59:59" → "2020-01-04 00:00:00"
- "1856-07-31 00:00:00" → "2020-01-01 00:00:00"

Another example with a minimum range of "2020-01-01 01:00:00" and a maximum range of "2020-01-01 03:00:00" with the unit set to **Hours** provides 3 possible mask values:

- "2020-01-01 01:00:00"
- "2020-01-01 02:00:00"
- "2020-01-01 03:00:00"

Using the same range of "2020-01-01 01:00:00" to "2020-01-01 03:00:00" but with the unit set to **Minutes**, there are 121 possible output values as the unit is the granularity at which time is subdivided. Note that the range is inclusive of both range values. Possible masked values may be as follows:

- "2020-01-01 01:00:00"
- "2020-01-01 01:14:00"
- "2020-01-01 01:59:00"
- "2020-01-01 02:23:00"
- "2020-01-01 03:00:00"

All inputs with the same value masked with the same algorithm configuration will result in the same output values.

Date Shift (Algorithm frameworks)

Extensible Algorithm Framework

The Date Shift framework masks date values to different dates based on a specified range around the input value. Masked values are calculated algorithmically using the algorithm's key, so rekeying the algorithm will cause different outputs to be generated for each input. All valid input values will be masked to a new value, and the new value will never match the input.

Creating a date shift algorithm via UI

The screenshot shows the 'Create Date Algorithm' interface. On the left, a 'Select Framework' sidebar lists various options, with 'Date' selected. The main panel is titled 'Create Date Algorithm' and contains the following elements:

- Algorithm Name:** A text input field.
- Description:** A larger text input area.
- Select Algorithm type:** A radio button selection with 'Shift' selected. A 'Learn More' link is next to it.
- Replacement:** Two input fields for 'Min Date' and 'Max Date', each with a calendar icon and a clock icon.
- Shift:** Two input fields for 'Min Value' and 'Max Value'.
- Unit:** A dropdown menu currently set to 'SECONDS'.
- Buttons:** 'Cancel' and 'Save' buttons at the bottom right.

1. In the upper right-hand region of the **Algorithm** tab under **Settings**, click **Add Algorithm**.
2. Select **Date**. The "Create Date Algorithm" pane appears.
3. Enter an **Algorithm Name**.
Info: This MUST be unique.
4. Enter a **Description**.
5. Under **Select Algorithm type** choose **Shift**.
6. Enter **Min Value** and **Max Value**. These values provide a range in which the masked value will differ from the input given a specified unit of time. The range is inclusive of both values where negative values represent units of time in the past and positive values represent units of time in the future. 0 may be included in the range or as one of the range values, but the input will not mask to the same value. A minimum value and maximum value that are equal will result in a fixed shift of that amount of time. For example, entering 3 as a min value and 3 as a max value with a unit of **Days** will mask all input values to 3 days in the future.
7. Check the **Roll** box to preserve all units of time larger and smaller than the specified unit. Only the value of the specified unit will change. This option is supported for units months, days, hours, minutes, and seconds.

8. Choose the **Unit** of time from the drop-down: **Years, Months, Days, Hours, Minutes, or Seconds**. This represents the unit of time the range is expressed in.
9. When you are finished, click **Save**.

For information on creating Date Shift algorithms through the API, see [API Calls for Creating Algorithms - Date Shift](#).

Examples

As an example, a Date Shift algorithm with a minimum value of 3 and a maximum value of 5 with the unit set to **Days** will shift the input value from 3 to 5 days into the future. Dates may mask as follows:

- "2021-02-03 12:30:00" → "2021-02-06 12:30:00"
- "1905-12-10 00:00:00" → "1905-12-15 00:00:00"
- "2001-07-31 23:45:30" → "2001-08-04 23:45:30"

With roll enabled and the same configuration, a date at the end of a month will wrap around to the beginning of the month. Dates may mask as follows:

- "2021-02-25 10:00:00" → "2021-02-01 10:00:00"
- "1932-05-03 01:15:15" → "1932-05-08 01:15:15"
- "1999-08-31 18:30:00" → "1999-08-03 18:30:00"

All inputs with the same value masked with the same algorithm configuration will result in the same output values.

Dependent Date Shift (Algorithm frameworks)

Extensible Algorithm Framework

The Dependent Date Shift algorithm provides a method to manipulate dates together where a dependency exists between the two dates that must be maintained. Examples of this include date of admission and date of discharge or date of birth and date of death. If we were to attempt to mask these dates independently, we may end up with a situation where a latter date such as date of discharge, was masked to be earlier than date of admission. If we were dealing with date of birth and date of death we may end up masking the values in a way that turns an 80 year old into a 5 month old. To this end, the Dependent Date Shift algorithm provides a way to mask these dependent dates in a way that:

- maintains the relationship between the dates (ie: the later date always stays later)
- maintains an approximate interval between the dates, within a provided `intervalRange / unit` combination

The Dependent Date Shift algorithm takes in 2 dates (designated `date1` and `date2`). It masks `date1` based on the provided values for `minRange`, `maxRange`, `unit` and `roll`. It then modifies the original interval based on `intervalRange` and `unit` to calculate `date2`. If the dates differ but the returned interval is zero (i.e.: the difference between the dates is smaller than the interval value), we assume the interval value to be 1 if `date2` is later than `date1` and -1 if `date1` is later than `date2`.

The masked results are deterministic for each pair of inputs with the same algorithm key and date and interval ranges. The algorithm does not allow for zero mask so all masked values will never be equal to the input. If `date1` is not provided, `date2` will be masked based on the provided values for `minRange`, `maxRange`, `unit` and `roll`.

Creating a dependent date shift algorithm via UI

1. In the upper right-hand region of the **Algorithm** tab under **Settings**, click **Add Algorithm**.
2. Select **DependentDateShift**. The "Create Dependent Date Shift Algorithm" pane appears.
3. Enter an **Algorithm Name**.
Info: This MUST be unique.
4. Enter a **Description**.
5. Enter a **Minimum Range**. This number represents the smallest number of time units that will be added to `date1` when masking. The range is inclusive of this value. Negative values represent units of time in the past and positive values represent units of time in the future. If `date1` is not provided, this is applied to `date2`.
6. Enter a **Maximum Range**. This number represents the largest number of time units that will be added to `date1` when masking. The range is inclusive of this value. Negative values represent units of time in the past and positive values represent units of time in the future. If `date1` is not provided, this is applied to `date2`.
7. Enter an **Interval Range**. A number representing the +/- range value to shift the interval inclusive of the range value. A value of 0 will not change the interval between dates. This number may not be less than 0. If the specified unit difference between `date1` and `date2` is within the bound of the `intervalRange`, only values

will be provided such that the sign of the difference is preserved. For example, if the day difference between `date1` and `date2` is 2 and the specified `intervalRange` is 3, only values greater than -2 will be used (i.e.: -1 to 3). Otherwise, the full range of values will be used (i.e.: -3 to 3).

8. Check the **Roll** box to preserve all units of time larger and smaller than the specified unit. Only the value of the specified unit will change. This option is supported for units months, days, hours, minutes, and seconds. This applies when masking `date1`. If `date1` is not provided, this is applied to `date2`.
9. Choose the **Unit** of time from the drop-down: **Years, Months, Days, Hours, Minutes, or Seconds**. This represents the unit of time the range is expressed in. This unit is also used to determine the interval between `date1` and `date2`.
10. When you are finished, click **Save**.

For information on creating Date Shift algorithms through the API, see [API Calls for Creating Algorithms - Dependent Date Shift](#).

Examples

As an example, a Dependent Date Shift algorithm with a minimum value of 3 and a maximum value of 5 and an interval Range of 5 with the unit set to **Days** will shift the `date1` input value by 3 to 5 days into the future. It will then change the interval by a range of +/-5 days from the original interval to mask `date2`. Dates may mask as follows:

- 1905-12-10 00:00:00, 1907-08-01 10:14:00 → 1905-12-13 00:00:00, 1907-08-06 00:00:00
- 2001-07-31 23:45:30, 2005-04-12 07:13:00 → 2001-08-03 23:45:30, 2005-04-12 23:45:30
- 2021-02-03 12:30:00, 2021-02-07 12:34:00 → 2021-02-06 12:30:00, 2021-02-14 12:30:00

With roll enabled and the same configuration, a date at the end of a month will wrap around to the beginning of the month. Dates may mask as follows:

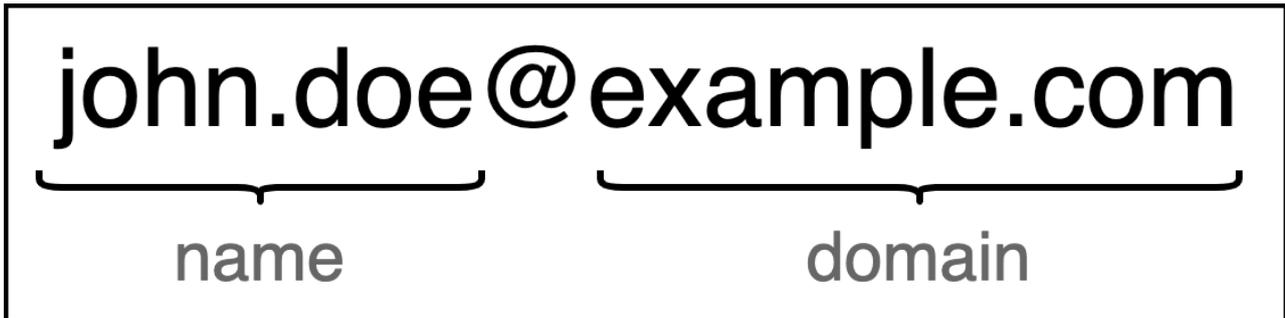
- 1905-12-10 00:00:00, 1907-08-01 10:14:00 → 1905-12-13 00:00:00, 1907-08-04 00:00:00
- 2001-07-31 23:45:30, 2005-04-12 07:13:00 → 2001-07-03 23:45:30, 2005-03-18 23:45:30
- 2021-02-03 12:30:00, 2021-02-07 12:34:00 → 2021-02-06 12:30:00, 2021-02-14 12:30:00

All inputs with the same value masked with the same algorithm configuration will result in the same output values.

Email (Algorithm frameworks)

Extensible Algorithm Framework

The Email framework masks string values by splitting the input on the '@' symbol and independently masking the name and domain portions of the email address. Masked values are calculated algorithmically using the algorithm's key, so rekeying the algorithm will cause different outputs to be generated for each input. All inputs to this framework are valid and the framework will not generate non-conformant data events. Note that it is possible for chained algorithms specified for the *Algorithm* option to generate non-conformant data events.



Malformed input handling

- Inputs without an '@' symbol: apply the name action to the entire input
- Inputs with no name portion: apply the domain action to the entire input
- Inputs with no domain portion: apply the name action to the entire input
- Inputs with no name portion and no domain portion: return an '@' symbol

Creating an email algorithm via UI

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email**
- Segment Mapping (legacy)
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Extended

Create Email Algorithm [Learn More](#)

Algorithm Name

Description

Mask Name With

Mask Domain With

Domain Replacement

1. In the upper right-hand region of the **Algorithm** tab under **Settings**, click **Add Algorithm**.
2. Select **Email**. The "Create Email Algorithm" pane appears.
3. Enter an **Algorithm Name**.
Info: This MUST be unique.
4. Enter a **Description**.
5. From the dropdown **Mask Name With**, choose one of the following options:
 - **Unique Value:** applies a SHA-256 hash of the entire input then Base32 encodes the hash value
 - **Lookup Value:** applies a secure lookup using the values provided in the uploaded file or file reference
 - **Algorithm:** applies the specified string type extensible algorithm**Info:**
The **Unique Value** option may produce masked name portions with lengths up to 52 characters.
6. If applicable, complete the configuration for masking the name portion as follows:
 - **Lookup Value:** upload a lookup file with new line separated values or provide a file reference
 - **Algorithm:** select a string type extensible algorithm to be used to mask the name portion of the input
7. From the dropdown **Mask Domain With**, choose one of the following options:
 - **Replacement Text:** replaces the domain portion with a fixed value
 - **Algorithm:** applies the specified extensible algorithm instance
8. Complete the configuration for masking the domain portion as follows:
 - **Replacement Text:** enter a value to replace the entire domain portion
 - **Algorithm:** applies the specified extensible algorithm instance
9. When you are finished, click **Save**.

For information on creating Email algorithms through the API, see [API Calls for Creating Algorithms - Email](#).

Examples

As an example, an Email algorithm that uses *Lookup Value* to mask the name portion and *Replacement Text* to mask the domain portion with the following configuration:

Lookup File:

```
Amy
Bob
Jake
Katherine
```

Replacement text: example.com

May mask as follows:

- "albert@delphix.com" → "Bob@example.com"
- "albert@gmail.com" → "Bob@example.com"
- "andrew_smith_123@delphix.com" → "Katherine@example.com"

Another example that uses the *Algorithm* option for both the name and domain portion with the following configuration:

Name algorithm: `dlpx-core:FirstName`

Domain algorithm: `dlpx-core:CM Alpha-Numeric`

May mask as follows:

- "bob@gmail.com" → "alton@dqpnx.fsy"
- "bob@hotmail.com" → "alton@poatzdw.bya"
- "alex@gmail.com" → "jameel@dqpnx.fsy"
- "joe_123@yahoo.com" → "miryam@wbpaq.kts"

 The Email framework will not generate non-conformant data events, but the chained algorithm may generate such events.

All inputs with the same value masked with the same algorithm configuration will result in the same output values.

Free Text Redaction (Algorithm frameworks)

Extensible Algorithm Framework

A Free Text Redaction Algorithm Framework helps you remove sensitive data that appears in free-text columns such as “Notes.” This type of algorithm requires some expertise to use because you must set it to recognize sensitive data within a block of text.

The algorithm uses a list of lookup words to determine what information it needs to mask. You can decide which words the algorithm uses to search for material such as addresses. For example, you can set the algorithm to look for “St,” “Cir,” “Blvd,” and other words that suggest an address. You can also use pattern matching to identify potentially sensitive information. For example, a number that takes the form 123-45-6789 is likely to be a Social Security Number. Lookup words and regular expressions will match individual words within the input text, rather than phrases.

You can use a Free Text Redaction Algorithm Framework to show or hide information by displaying either a “DenyList” or an “AllowList.”

DenyList – Designated material will be redacted (removed). For example, you can set a deny list to hide patient names and addresses. The deny list feature will match the data in the lookup file to the input.

AllowList – ONLY designated material will be visible. For example, if a drug company wants to assess how often a particular drug is being prescribed, you can use an allow list so that only the name of the drug will appear in the notes.

Creating a free text redaction algorithm via UI

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping (legacy)
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Extended

Create Free Text Redaction Algorithm

Algorithm Name

Description

Redact Type

DenyList ▼

Lookup File - Upload or specify a URI for text file

Select...

Lookup File Redact Value

Replacement Value for LookupFile

Regular Expressions - Specify new line separated java 8 style RegEx

Regular Expression Redact Value

Replacement value for RegEx

Cancel

Save

1. Enter an **Algorithm Name**.
2. Enter a **Description**.
3. Select a **Redact Type**: the **Deny List** or **Allow List**.
4. Select a **Lookup File** and enter a **Redaction Value** OR/AND
5. Enter **Regular Expressions** separated by a new line and enter a **Redaction Value**.
6. Click **Save**.

Existing limitations:

1. The maximum number of supported **Regular Expressions** is **50**. Exceeding this number will lead to the Component Configuration exception.
2. The maximum number of supported words in the **Lookup File** is **1000**. Exceeding this number may affect the algorithm performance.
3. The **Lookup File** format must be **txt**.
4. Every entry in the **Lookup File** must be a new line separated. Phrases are not supported. Case sensitive.
5. The maximum length of an input text to mask is **32768**. Exceeding this number will lead to the Non-Conformant data exception.

For information on creating Free Text Redaction algorithms through the API, see [API Calls for Creating Algorithms - Free Text Redaction](#).

Examples

Input:

```
The customer Bob Jones is satisfied with the terms of the sales
agreement. Please call to confirm at 718-223-7896.
```

Algorithm configuration:

1. The Redact Type is **DenyList**
2. Lookup File entries:

```
Bob
Jones
agreement
```

3. The Lookup File Redaction Value is **XXXX**
4. Regular Expressions entry:

```
[0-9]{3}-[0-9]{3}-[0-9]{4}
```

- a. The Regular Expression Redaction Value is **YYYY**

Masking result:

```
The customer XXXX XXXX is satisfied with the terms
of the sales XXXX. Please call to confirm at YYYY.
```

"Bob", "Jones", "agreement" and the phone number are redacted.

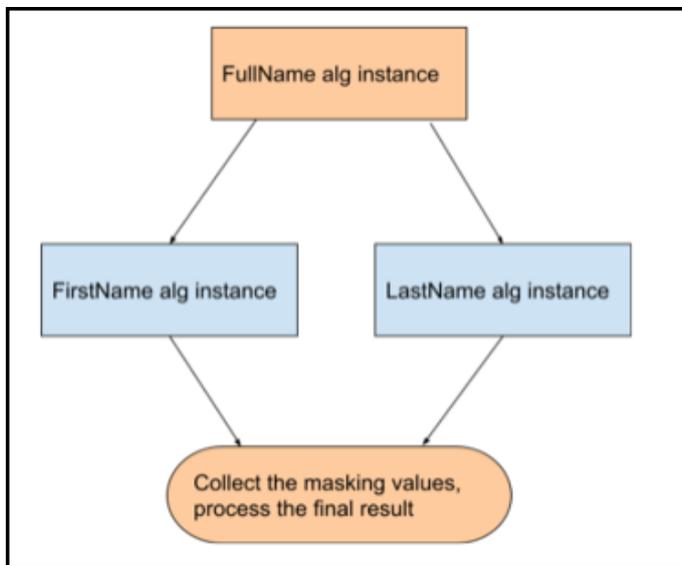
Full Name (Algorithm frameworks)

Extensible Algorithm Framework

The Full Name algorithm (introduced in Masking Engine version 6.0.8.0) has logic to separate the input into two parts: First and Last names. It can also limit the number of masked names (removing the rest) and “smart trim” the result (masked) output to the required length.

If Name framework algorithm instances are used, the Full Name algorithm can consider particles. The Full Name algorithm uses the Last Name algorithm’s particles when determining which part of the input is the last name. The remainder of the input is considered the first name and the First Name algorithm’s particles apply to that.

After distinguishing the parts of the input string, the Full Name algorithm feeds each word from the first name part (which also includes middle names, treated the same as first names) individually to the First Name algorithm instance, and the whole last name part to the Last Name algorithm instance. Then it combines the masking results, according the embedded logic and the configuration.



If the input string contains only a single word - this word is considered as a first name or last name (depending on the **Consider Single Word Input as Last Name** flag) and forwarded for masking to the corresponding chained algorithm instance. A single word input is always masked, even if it is a configured particle.

Main features of the Full Name Framework:

- Deterministic output: The masked result for each input is consistent when using the same algorithm key, same configuration and same chained algorithm instances.
- Not unique: The masked result might be the same for different inputs.
- Garbage in garbage out: the algorithm returns the unmasked input / null / empty string if input is one of the following: null, empty string “”, white spaces only “ ”.
- Single word input: considered either as a Last Name (default) or as a First Name, even if configured in one of the particles files.
- When particle is configured in both particles files: the remove action takes precedence.
- Multiple first names: masks only first N names (1-4, as configured, default = 2), the rest are removed. (Note: only one name can be considered the last name; the rest are masked as first names.)
- Full Name Convention: if a configured last name separator is detected, or the configured convention is “last-first-middle”, then the input is interpreted as last-first-middle. Otherwise, it is first-middle-last (default). Leading/trailing and duplicate white spaces are not preserved.

- Smart trim: if trimming of the masked value is required, it's done in a way to keep the full name as long and realistic looking as possible. For instance, first we trim the leading/trailing preserved particles. If that's not enough, we abbreviate the masked first/middle names (one by one, starting with the last one). If that's still not enough, we remove the particles prior to the last name, etc.

Below is an example of smart trim. Let's suppose our masked result (prior to checking of the maxLength) is:

"President George Herbert Walker Van Bush Jr."		
maxLengthOfMaskedFullName value	action	result
55	Nothing. The string is shorter.	President George Herbert Walker Van Bush Jr.
42	Cut particle at the end (if known)	President George Herbert Walker Van Bush
30	Cut particle at the beginning (if known)	George Herbert Walker Van Bush
26	Abbreviate FNs starting last	George Herbert W. Van Bush
22	Continue abbreviate FNs	George H. W. Van Bush
17	Continue abbreviate FNs	G. H. W. Van Bush
14	Cut FN abbreviation(s) starting last	G. H. Van Bush
11	Continue cutting abbreviations	G. Van Bush
8	Leave LN if possible	Van Bush
4	Leave LN if possible	Bush
2	Cut the LN from the end	Bu

The chained instances for First Name and Last name masking can be any existing extensible algorithm instance that masks the String type. However, it is recommended to use the instances based on the Name framework. Particles can only be considered for Name framework algorithms, and it's recommended that the First Name and Last Name algorithm instances have the same particle files.

Creating a Full Name algorithm via UI

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Numeric Expression
- IBAN
- Extended

Create Full Name Algorithm [Learn More](#)

Algorithm Name

Description

Mask Type

STRING

First Name Algorithm	Last Name Algorithm
<input type="text" value="dlpx-core:FirstName"/>	<input type="text" value="dlpx-core:LastName"/>

Maximum First Names	Maximum Masked Full Name Length
<input type="text" value="2"/>	<input type="text" value="0"/>

Full Name Convention

Consider Single Word Input as Last Name

Last Name Separators Separated by comma(,)

 Include comma

1. In the upper right-hand corner of the **Algorithm** tab, click **Add Algorithm**.
2. Select the **Full Name** Framework. The **Create Full Name Algorithm** pane appears.
3. Enter an **Algorithm Name**. (Required)

Info:
This MUST be unique on the Masking Engine.
4. Enter a **Description**. (Optional)
5. Choose the **First Name Algorithm**. (Required) In the dropdown menu you will be suggested to choose from the existing extensible algorithms of String type.
6. Choose the **Last Name Algorithm**. (Required) In the dropdown menu you will be suggested to choose from the existing extensible algorithms of String type.
7. Choose the **Maximum First Names** configuration. (Optional. Integer. min value = 1, max value = 4, default = 2) Total number of first/middle names to be masked. The rest would be ignored.
8. Choose the **Maximum Masked Full Name Length**. (Optional. Integer. Default is 0) This number should be ≥ 0 (i.e. not negative). This is the maximum number of characters for the masked output string. The masked

result is “smart trimmed” to fit that length (see explanation and example table above for details on smart trimming). Value 0 means length is unlimited.

Info:

We also try to detect the length of the destination field. Some Data Sources provide that value, while others don't. For example: if Data Source provides value **10** for the destination column length and current configuration field is set to **0** or any value longer than 10 - the shortest value wins, i.e. in this example masked result would be trimmed to 10 characters.

- Specify a **Full Name Convention**. (Optional. Enum. Default: "First-Middle-Last") Dropdown menu provides choice of 2 values:

First-Middle-Last
Last-First-Middle

This configurations helps the Full Name algorithm distinguish between first name(s) and last name, if **Last Name Separator(s)** are not configured or not detected in the input string.

- Choose the **Consider Single Word Input as Last Name**. (Optional. Boolean. Default is true) If chosen (default case) - consider the single word input as a last name. Otherwise as a first name.
- Configure **Last Name Separators** (Optional. List. Default: contains comma ',') Here you can specify comma separated single punctuation marks (but hyphen '-' and dot '.', which are reserved for another logic) which will serve for identifying the last name in the input. First identified separator makes that distinguishing, rest are ignored. To choose comma ',', there is a separate field aside **Include comma**. By default comma is included as a separator.

Here is an example of how last name separator works:

Let's suppose our configured separators are comma ',' and colon ':':

Input: "dela Cruz, Maria Cristina: Manansala"

The first detected separator (framework reads the input left to right) is after word "Cruz".

So "dela Cruz" will be detected as a last name part, and "Maria Cristina: Manansala" as a first names.

Masking result would be in the same order with the same separator, for example: "Maritnas, Antonio Stephan".

- When you are finished, click **Save**.

For the description of any configurable field you can open a popup window by pressing on the blue "? Learn More" link in the upper right corner:

Full Name Algorithm

This is the framework to cover the scenarios where it is required to mask a full name input string with deterministic and not unique masking results. The goal of a Full name masking framework is to mask the first, middle and last names consistently when they appear in the same field using algorithms that can be applied to the component parts (e.g. First Name alone.).

Framework Options:

First Name Algorithm (Required): String type Extensible Algorithm instance to be used to mask first and middle name of the full name input.

Last Name Algorithm (Required): String type Extensible Algorithm instance to be used to mask last name of the full name input.

Maximum First Names: Total number of first/middle names to be masked, the rest would be ignored. Minimum value is 1, maximum is 4 and **default** is 2;

Maximum Masked Full Name Length: Max number of characters or length of masked output string. **default** is 0, which means unlimited.

Full Name Convention: A flag to configure the input full name convention of pattern

- **First-Middle-Last (Default)** - The input name starts with a first name.
- **Last-First-Middle** - The input name starts with a Last name.

Consider Single Word Input as Last Name: A flag to configure if single word input should be masked with Last name or not. **default** is on/true.

Last Name Separators: A comma separated list of characters which should be considered as first and last name separators.

Include comma: A flag to include "," as last name separator character.

OK

For information on creating Full Name algorithms through the API, see [API Calls for Creating Algorithms - Full Name](#).

Mapping (Algorithm frameworks)

Extensible Algorithm Framework

A Mapping algorithm allows you to state what values will replace the original data. It maps original data values to masked values that are pre-populated to a lookup table through the Masking Engine user interface. There will be no collisions in the masked data because it always matches the same input to the same output. For example “David” will always become “Ragu,” and “Melissa” will always become “Jasmine.” The algorithm checks whether an input has already been mapped; if so, the algorithm changes the data to its designated output.

You can use a Mapping algorithm on any set of values, of any length, but you must know how many values you plan to mask. You must supply AT MINIMUM the same number of values as the number of unique values you are masking; more is acceptable. For example, if there are 10,000 unique values in the column you are masking you must give the Mapping algorithm AT LEAST 10,000 values.

The Mapping Algorithm can be configured for mappings managed locally on the Masking Engine or remotely on a customer managed PostgreSQL database. The remote configuration should be used if the customer wishes to more easily manage the storage allocated for mappings, or if there is a desire to share the same Mapping Algorithm mappings across multiple Masking Engines. More information about remote mapping configuration can be found [here](#).

Masking Engine 6.0.9.0 and earlier: When you use a Mapping algorithm, you cannot mask more than one table at a time. You must mask tables serially.
Masking Engine 6.0.10.0 and later: A single Mapping Algorithm can have multiple jobs running concurrently.

Tokenization/Reidentification

Given the nature of Mapping Algorithms, they can be used with Tokenization and Reidentification jobs. However, if `ignoreCharacters` are configured for the algorithm, Tokenization/Reidentification cannot be used.

Sync

Mapping Algorithm can be synced in 1 of 2 ways:

1. **Syncing a locally managed Mapping Algorithm:** This can be done to effectively *make a copy* of an algorithm from one Masking Engine to another. In addition to syncing the algorithm, the mappings must be manually exported from the source engine and imported into the target engine. Once this is complete, the 2 algorithms (on the source and target) will have the same names and initial set of mappings (at the time of sync) but will function as 2 separate algorithms. That is to say, adding new mappings on the source *will not* have any impact on the algorithm on the target.
2. **Syncing a remotely managed Mapping Algorithm:** This can be done to *share* the same Mapping Algorithm across Masking Engines. In this case, once synced, the algorithm on the source and target(s) would point to the SAME remote mapping database. This would mean that adding/removing/manipulating the mappings would affect the algorithm on all engines that use it.

For more information on sync, see [here](#).

Creating a mapping algorithm via UI

1. In the upper right-hand corner of the **Algorithm** tab, click **Add Algorithm**.
2. Select **Mapping**.
3. The **Create Mapping Algorithm** pane appears.
4. Enter an **Algorithm Name**. **Info:** This MUST be unique.
5. Enter a **Description**.

- Select whether or not the mappings will live locally or remotely, by toggling the **Local Mapping Store** checkbox appropriately. If using a local mapping store, proceed to step 9. Info: For more information about remote mapping stores, click [here](#).

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping (legacy)
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Extended

Create Mapping Algorithm

Algorithm Name

Description

Local Mapping Store

Ignore Characters Separated by comma(,)

Ignore comma(,)

Manage Mappings
Cancel
Save

- Specify **Host/IP**, **Port**, **Mapping Database**, and **Schema** of the remote database.

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping (legacy)
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Extended

Create Mapping Algorithm

Algorithm Name

Description

Local Mapping Store

Host/IP **Port**

Mapping Database **Schema**

Mapping Connection Properties

Select...

Ignore Characters Separated by comma(,)

Ignore comma(,)

Manage Mappings

Cancel

Save

8. Enter any remaining connection parameters in a properties file specified by the **Mapping Connection Properties** field.
9. To ignore specific characters, enter one or more characters in the **Ignore Character List** box. Separate values with a comma.
10. To ignore the comma character (,), select the **Ignore comma (,)** checkbox.
11. When you are finished, click **Save**.

Before you can use the algorithm by specifying it in a profiling job, you must add it to a domain. If you are not using the Masking Engine Profiler to create your inventory, you do not need to associate the algorithm with a domain.

For information on creating Mapping algorithms through the API, see [API Calls for Managing Algorithms - Mapping](#).

Managing mappings via UI

Regardless of where the mappings reside (local or remote), the management process is the same.

To start go to the **Edit Mapping Algorithm** screen and select **Manage Mappings**

At the top there are 2 statistics provided for the mappings:

1. **Total Mappings** is the number of mapping outputs that exist for this algorithm.
2. **Available Mappings** is the number of mappings that have not yet been assigned to an input value.

- i** 1. When a job using the Mapping Algorithm runs, the mappings are loaded into memory. This means that enough memory must be provided to the job to load the mappings. A Mapping Algorithm with 2GB worth of mappings will require a job with a larger configured XMX than what is needed for a Mapping Algorithm with 2MB worth of mappings.

In addition to the mapping statistics there are 4 actions to choose for managing mappings:

Delete mappings

This action will delete all input/output combinations and effectively start this algorithm fresh. For this option to take effect you must select the **Delete Mappings** action and then click **Delete**.

Manage Mappings

Total Mappings: 0
Available Mappings: 0

Action

Delete Mappings ▼

DeleteBack

Export mappings

This action will export all mappings into a file that can then be used to seed another mapping algorithm or to simply have a list of established mappings. For security purposes a passphrase is required to encrypt the file on export.

To export mappings select the **Export Mappings** action and provide a **passphrase** and then click **Export**.

Once the export file has been generated a link that says **Click here to Download File** will appear. Click this to download the export file.

- i** If you wish to decrypt the exported file from the command line, run the following command:
`openssl enc -aes-128-cbc -a -d -pass stdin -pbkdf2 -iter 100000 -md SHA256 -in PATH_TO_EXPORT_FILE`

Manage Mappings

Total Mappings: 0

Available Mappings: 0

Action

Export Mappings ▼

Passphrase

Enter Passphrase for the Mapping file

Export

Back

Import mappings

This action will add mappings to the mapping algorithm. Mappings can be provided in 2 different formats - **PLAINTEXT** and **CSV**.

Manage Mappings

Total Mappings: 0

Available Mappings: 0

Action

Import Mappings ▼

File Type

CSV ▼

Import Mappings/Outputs - Upload Mapping File

Select...

Passphrase

Enter Passphrase for the Mapping file

Import

Back

PLAINTEXT

A PLAINTEXT mapping file can ONLY provide mapping outputs (i.e.: values you want to mask to). The file must have NO header. Make sure there are no spaces or returns at the end of the last line in the file. The following is sample file content. Notice that there is no header and only a list of values.

```
Smallville
  Clarkville
  Farmville
  Townville
  Cityname
  Citytown
  Towneaster
```

CSV

A CSV mapping file can provide both mapping inputs and outputs. That is, you can determine beforehand what you want your mappings to be. The CSV file MUST have ONLY 2 columns - input and output. The first line of the file MUST be the header "input,output". The following is a sample CSV mapping file.

```
input,output
New York,Smallville
Boston, Clarkville
San Francisco, Towville
",Cityname
",Citytown
",Toweaster
```

i You may opt not to specify an input, but you must specify an output for a line to be considered valid. Invalid lines are silently ignored.

Once a **File Type** is selected, choose the mapping file in the **Import Mappings/Outputs** field.

i If providing a previously exported mapping file which has been encrypted with a passphrase, select the **CSV** file type, provide the *unaltered* encrypted file and provide a **passphrase**.

When the appropriate selections have been made, click **Import**.

i Any duplicate values provided will be silently ignored.

Reset mappings

This action will delete all inputs for provided mappings, giving you a mapping algorithm with as many outputs as you had before, but with all of them available for assignment the next time the mapping algorithm is used.

Manage Mappings

Total Mappings: 0
Available Mappings: 0

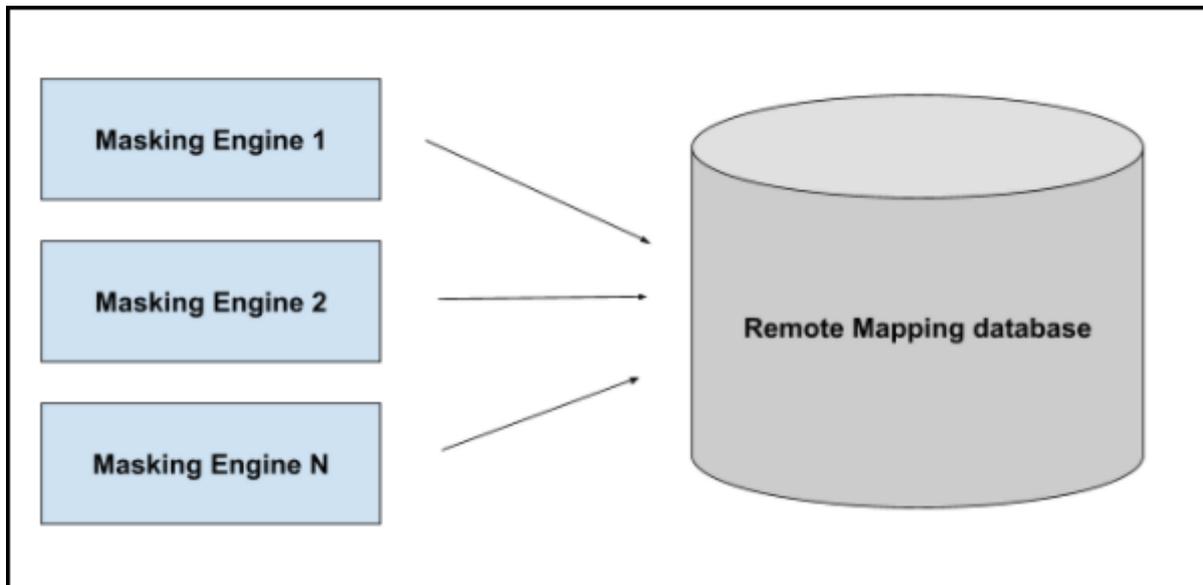
Action

Reset Mappings ▼

ResetBack

Remote Mapping

With the release of version 6.0.10.0 of the Masking Engine, the Mapping Algorithm now provides support for storing all mappings on a user-supplied database. This enables users to share mappings for the same Mapping Algorithm across engines. The mapping database connection info can be provided when a Mapping Algorithm is added or edited.



In order to serve as a mapping database, the following requirements must be met:

- The database must be a PostgreSQL database version 9.5 or newer.
- The database must be reachable by the Masking Engine

All necessary tables and functions to successfully run the Mapping Algorithm will be created by the Masking Engine upon connection to the remote mapping database.

Remote mappings are managed in the same way as local mappings via the Masking Engine GUI or APIs.

i It is completely fine to use the same remote database for multiple Mapping Algorithms on the same Masking Engine or across many Masking Engines.

i Given that the Masking Engine will need to query the remote database, network latency will have an effect on how fast a job running a Mapping Algorithm will run, especially on the "initial" run of a Mapping Algorithm when the majority of new mappings are established.

Expectations

By opting to manage their own mappings, the user agrees to be responsible for:

- Database uptime
- Database security
- Network connectivity
- Database storage

Configuring the connection

The user may opt to configure their PostgreSQL database however they wish. With the exception of host, port, database and schema, all other connection properties may be provided via a properties file, per the [PostgreSQL JDBC Driver documentation](#).

For databases with SSL/TLS connections, the correct properties should be supplied via the properties file.

Min Max (Algorithm frameworks)

Extensible Algorithm Framework

The Continuous Compliance Engine provides two Min Max algorithm frameworks: "MinMax Date" and "MinMax Number" to normalize data within a range. Values that are extremely high or low in certain categories allow viewers to infer someone's identity, even if their name has been masked. For example, a salary of \$1 suggests a company's CEO, and some age ranges suggest higher insurance risk. You can use a Min Max algorithm to move all values of this kind into the midrange. This algorithm allows you to make sure that all the values in the database are within a specified range. The algorithm frameworks are applicable to numeric or date data types.

The **Replacement Value for Nonconforming Data** value is used when the underlying data to be masked is of type String and conversion to a date or a number is required.

Creating a Min Max Algorithm via UI

1. Enter the **Algorithm Name**.
Info: This MUST be unique.
2. Enter the **Description**.
3. Enter the **Min Number** and the **Max Number**.
4. Enter the **Replacement Value for Nonconforming Data** if needed.
5. Click **Save**.

For information on creating Min Max algorithms through the API, see [API Calls for Creating Algorithms - Min Max](#).

Examples

Example: Age less than 18 years - enter Min Number 0 and Max Number 18.

Name (Algorithm frameworks)

Extensible Algorithm Framework

Starting in version 6.0.8.0, Delphix has introduced a builtin Extensible *Name* Algorithm Framework, co-existing with the legacy *FIRST NAME SL* and *LAST NAME SL* ones. Name Framework provides masking functionality for String type input. It's based on Secure Lookup mechanism, and includes additional configuration flags making it more flexible and robust.

Similar to Secure Lookup it creates masking results which are deterministic (i.e. the same algorithm with the same configuration and security key will provide the same result for the same input) and not unique. If you are looking for a framework whose algorithm(s) will provide unique masking results, you should consider using other frameworks (for example Character Mapping).

The new framework uses SHA256 hashing method and allows case configurations for input and output (i.e. masked) values. It also allows filtering accents and configuring the maximum length of the masked value and the maximum number of masked names returned. A multi-word input name may contain particles, such as prefixes, suffixes, titles, etc. The new framework allows configuring which particles are to be preserved and removed.

Creating a Name Algorithm via UI

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Numeric Expression
- IBAN
- Extended

Create Name Algorithm [Learn More](#)

Algorithm Name

Description

Mask Type

STRING

Case Sensitive Lookup

Filter Accent

Maximum Number of Names

Maximum Masked Name Length

Output (Masked) Case

Preserve Input Case

Particles to Preserve - Upload a File or Specify a URI

Select...

Particles to Remove - Upload a File or Specify a URI

Select...

Lookup - Upload a file or Specify a URI

Select...

1. In the upper right-hand corner of the **Algorithm** tab, click **Add Algorithm**.
2. Select the **Name** Framework. The **Create Name Algorithm** pane appears.
3. Enter an **Algorithm Name**. (Required)

Info:

This MUST be unique on the Masking Engine.

4. Enter a **Description**. (Optional)
5. Choose the **Case Sensitive Lookup** configuration. The default is *false (unchecked)*, which means lookup is not case sensitive, so the same input of different cases will be masked to the same value. For example:

```
Peter -> John
peter -> john
```

If the **Case Sensitive Lookup** box is checked, then the same input of different cases will be masked to the different values, for example:

```
Peter -> John
peter -> andrew
```

6. Choose the **Filter Accent** configuration. The default is *true (checked)*. If the **Filter Accent** box is checked, then similar input with and without accented symbols will be masked to the same values. For example:

```
Adrián -> John
Adrian -> John
```

If the **Filter Accent** box is unchecked, it will be masked to the different values, for example:

```
Adrián -> John
Adrian -> Peter
```

7. Choose the **Maximum Number of Names**. (Range 1-4, default 2). This is the maximum number of names to be masked and returned. The rest are dropped.
8. Choose the **Maximum Masked Name Length**. (A number greater than or equal to 0, default is 0). This is the maximum number of characters or length of the masked output string. The masked result is trimmed to fit that length. Value 0 means length is unlimited.

Info:

We also try to detect the length of the destination field. Some Data Sources provide that value, while others don't. For example: if Data Source provides value **10** for the destination column length and current configuration field is set to **0** or any value longer than 10 - the shortest value wins, i.e. in this example masked result would be trimmed to 10 characters.

Warning:

Some UTF-8 characters might take 2 bytes. If lookup file contains those characters - the trimmed result might be not as expected, since we trim by the number of characters and not number of bytes. There is a bug open for that mismatch.

9. Choose the **Output (Masked) Case** configuration. (Default is *Preserve Input Case*)

Output (Masked) case options:

- **Preserve Lookup File Case** - keep masked value as found in the Lookup File.
- **Preserve Input Case (Default)** - check the input case, which can be one of the following three:
 - All uppercase - in that case force whole masked value to uppercase
 - All lowercase - in that case force whole masked value to lowercase
 - Mixed (if at least 1 character case is different from others) - in that case keep masked value as found in the Lookup File
- **Force all Uppercase** - forces whole masked value to uppercase
- **Force all lowercase** - forces whole masked value to lowercase

10. Specify a **Particles to Preserve** File. (Optional. A locally chosen file or a FileReference) Contains a list of particles to be preserved. These particles are not masked. For example, if the file contains the particle "von":

```
von Froum -> von Smith
```

11. Specify a **Particles to Remove** File. (Optional. A locally chosen file or a FileReference) Contains a list particles to be removed. These particles are removed prior to masking and do not affect masking result. For example, if the file contains the particle "von":

```
von Froum -> Smith  
Froum -> Smith
```

Info:

If particle is found in both "Preserve" and "Remove" files - it will be removed.

If the input contains only particles, a particle will be masked as if it were a name.

12. Specify a **Lookup** File. (Required. A locally chosen file or a FileReference)

This file is a single list of values. It does not require a header. Every line of the Lookup File might be used as a masked value. The Lookup File must be ASCII or UTF-8 encoding compatible. The following is sample file content:

```
Ann  
Marie  
Tomas  
Ann-Marie  
Basil  
Mark
```

13. When you are finished, click **Save**.

For information on creating Name algorithms through the API, see [API Calls for Creating Algorithms - Name](#).

Numeric Expression (Algorithm frameworks)

Extensible Algorithm Framework

Numeric Expression algorithms mask numeric input by evaluating it within a one-line, mathematical expression written by the user in the Java programming language. The expression can reference the current unmasked value via an implicit variable called `input`.

For example, to mask a numeric column by always multiplying the input by 50%, the following expression could be used:

```
input * 0.5
```

In addition to `input`, the expression can reference user-defined constant variables whose values are determined at the beginning of a masking job and remain fixed for the life of the masking job.

See below for examples of expressions and constants.

Creating a numeric expression algorithm via UI

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Numeric Expression**
- Extended

Create Numeric Expression Algorithm [Learn More](#)

Algorithm Name

Description

Expression

Input Type

double ▼

Replacement Value for Nonconforming Data

Constants [Learn More](#)

Name	Value	
Java variable	Java expression	Add

Cancel

Save

1. In the upper right-hand corner of the **Algorithms** tab, click **Add Algorithm**.

2. Select **Numeric Expression** Framework. The **Create Numeric Expression Algorithm** pane appears.
3. Enter an **Algorithm Name**. (Required)

Info

This MUST be unique on the Masking Engine.

4. Enter an optional **Description**.
5. Enter an **Expression**. This must be a one-line, mathematical expression written in the Java programming language that references `input` (the current unmasked value), e.g. `input * 0.5` or `input + Math.random()`. See below for more examples of expressions.
6. Choose the **Input Type**. This is the data type that `input` conforms to within the expression. The default `double` option causes `input` to be treated as a double-precision floating point variable in expressions such as:

```
input * 0.5
```

or

```
input + Math.random()
```

Input Type can also be set to `long`, which causes `input` to be treated as a long integer variable in expressions such as:

```
Long.sum(input, 50L)
```

The final **Input Type** option is `BigDecimal`, which causes `input` to be treated as a `java.math.BigDecimal` variable in expressions such as:

```
input.scaleByPowerOfTen(3)
```

7. Enter an optional **Replacement Value for Nonconforming Data** if necessary. This is the default masked value to be used if the unmasked input is not a numeric data type and can't automatically be converted to one.
8. Optional: define any constants used by the expression. Constants are variables that the expression can reference by name and whose values remain fixed for the life of a masking job. For example, to mask every column value in a masking job by multiplying them all by the same random number, you could use an expression such as:

```
input * theSameRandomNumber
```

but `theSameRandomNumber` would need to be defined as a constant whose **Name** is

`theSameRandomNumber` and whose **Value** is something like `new`

`java.util.Random().nextDouble()`. See below for more examples of constants.

9. When you are finished, click **Save**.

For information on creating Numeric Expression algorithms through the API, see [API Calls for Creating Algorithms - Numeric Expression](#).

Writing good expressions & constants

Expressions and the Java programming language are powerful. Care must be taken to avoid writing bad expressions, which will manifest in the form of failed masking jobs. It is highly recommended to stage complex expressions with a Java IDE such as [Eclipse](#) or [IntelliJ IDEA](#) before using them in a masking job.

The requirement that expressions must be written in Java might be intimidating to non-programmers, but simple mathematical equations in Java look similar to simple mathematical equations in general. The four most common operators are supported: addition (`+`), subtraction (`-`), multiplication (`*`), and division (`/`). For operators not supported by Java, use methods from the [java.lang.Math](#) library. For example, one might expect `input ^ 5` to mean "take input to the fifth power," but `^` is not a power operator in Java. Instead, use `Math.pow(input, 5.0)`.

To isolate parts of the expression for clarity or to enforce order of operations, use open and closed parentheses (`()`) only. Do not use square braces (`[]`) or curly braces (`{}`).

Expression do's and don'ts

Do use an **Input Type** (explained above) that corresponds to the data type of the column being masked. For columns whose values are floating-point numbers (i.e. numbers that have digits to the right of the decimal point) set **Input Type** to *double* (the default) or *BigDecimal* if the expression needs to treat the input as a [java.math.BigDecimal](#) object in order to perform more complex math. For columns whose values are integers (whole numbers), set **Input Type** to *long*.

Don't write expressions that do mathematically impossible things (e.g. divide by zero) or will result in numeric overflow or values that are too large or too small to fit in the database column being masked.

Don't use line breaks or other whitespace to force an expression to be longer than one line.

Don't attempt to assign an expression to a variable. For example, this won't work:

```
output = input * 0.5
```

but this will:

```
input * 0.5
```

The result of the expression will be automatically assigned as the masked value. It's not necessary or allowed to assign it to anything else.

Don't use the `return` keyword or end the expression with a semicolon.

Don't write expressions that return a non-numeric value, e.g.

```
java.util.Arrays.asList(input)
```

The above expression would return a `List` object, which can't be converted into a numeric value. Expressions must return a value whose type is numeric: an `int`, `short`, `long`, `float`, or `double` Java primitive type (or their object wrappers) as well as `java.math.BigDecimal` and `java.math.BigInteger`. Returning

`String` and `char[]` (character array) values is also acceptable as long as they can be converted into a numeric value.

Do fully-qualify any Java class the expression references that isn't in the `java.lang` package, e.g.

```
input * new java.util.Random().nextDouble()
```

This won't work:

```
input * new Random().nextDouble()
```

because Java's `Random` class is in the `java.util` package rather than `java.lang`.

Don't use the `import` keyword in an attempt to import non-`java.lang` classes that are referenced frequently by the expression and/or constants. Fully-qualify such Java classes every time they're referenced.

Constants

Constants are variables that the expression can reference by name and whose values remain fixed for the life of a masking job. Constant names must be [valid Java variable names](#). No two constants can have the same name, nor can "input" or "seed" be used as a constant name.

Constant values are very much like the expression: one-line Java expressions that must return a value. However, unlike the algorithm's main expression, constant values aren't required to be numeric.

Constants can reference by name other constants defined before them.

seed

There is a built-in constant named `seed`. Its value is a long integer that's based on the algorithm key, so the value of `seed` is guaranteed to remain the same across multiple masking jobs as long as the algorithm key remains the same. A common use case for `seed` is to seed a random number generator to produce the same (i.e. predictable) "random" number(s) among different masking jobs.

Numeric Expression Examples

Example 1

A numeric column must be masked by multiplying all of its values by the same random percentage. The random percentage must remain the same across every masking job.

Solution:

A single constant is required for the random percentage:

Name	Value
<code>randomPercentage</code>	<code>new java.util.Random(seed).nextDouble()</code>

Note that the built-in `seed` constant is being used to seed the random number generator, an instance of `java.util.Random`, which is used to produce a single random number.

The expression can then reference `randomPercentage` like this:

```
input * randomPercentage
```

Example 2

A numeric column must be masked by taking the square root of each value, then rounding it to a certain number of decimal places. Initially, it will be rounded to two decimal places, but the number of decimal places will be changed frequently, so it should be easily adjustable by the user.

Solution:

We'll define two constants this time:

Name	Value
<code>decimalPlaces</code>	<code>2</code>
<code>multiplier</code>	<code>Math.pow(10.0, decimalPlaces)</code>

then use this expression:

```
Math.floor(Math.sqrt(input) * multiplier + 0.5) / multiplier
```

The heavy lifting is being done by the main expression, which uses the `multiplier` constant. Note that `multiplier` references `decimalPlaces`, whose value could be easily changed by someone who is not inclined mathematically and doesn't understand how the expression is rounding numbers.

Example 3

We must mask a numeric column that represents the day of the current month, e.g. 1-31 (or 1-28, 1-29, 1-30). This column will be masked by adding to it a random number of days, which can be between 1 and the highest day in the current month, inclusive. If the masked value exceeds the highest day in the current month, it will simply be set to the highest day in the current month.

Solution:

First, since the day of the current month is an integer (whole number), set the algorithm's **Input Type** to *long* (integer) instead of the default *double* (floating point).

Then define three constants:

Name	Value
<code>calendar</code>	<code>java.util.Calendar.getInstance()</code>
<code>lastDayOfMonth</code>	<code>calendar.getActualMaximum(java.util.Calendar.DAY_OF_MONTH)</code>

Name	Value
randomDays	<code>new java.util.Random().ints(1, lastDayOfMonth + 1).iterator().nextInt()</code>

`calendar` is a new instance of `java.util.Calendar` set to the current date and time.

`lastDayOfMonth` uses `calendar` to determine the last day of the current month.

`randomDays` uses `lastDayOfMonth` to generate a random number between 1 and `lastDayOfMonth` (inclusive).

The expression will then look like this:

```
(input + randomDays > lastDayOfMonth) ? lastDayOfMonth.longValue() : input + randomDays
```

This expression leverages Java's ternary operator to mask conditionally. If the unmasked input plus `randomDays` exceeds `lastDayOfMonth`, then the masked value will simply be `lastDayOfMonth`. Otherwise, the masked value will be the unmasked input plus `randomDays`.

Payment Card (Algorithm frameworks)

Extensible Algorithm Framework

The Payment Card framework masks payment card numbers based on the starting digits to be preserved and the minimum number of positions to be masked. This framework is built on top of the [Character Mapping Algorithm Framework](#) with a character set of [0-9]. All characters outside of this character group remain unmasked. Masked values are calculated algorithmically using the algorithm's key, so rekeying the algorithm will cause different outputs to be generated for each input. The last digit may remain the same if the calculated check digit is equivalent to the last digit of the input. Any inputs with more than one digit will never mask to the original value.

⚠️ Any inputs with a single digit will remain unmasked.

This framework preserves the validity of the payment card number using the Luhn check. All input values with valid Luhn checks will be masked to values with valid Luhn checks. All invalid values with invalid Luhn checks will be masked to values with invalid Luhn checks.

Creating a payment algorithm via UI

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card**
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping (legacy)
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Extended

Create Payment Card Algorithm

Algorithm Name

Description

Minimum Masked Positions

Preserve Starting Digits

1. In the upper right-hand region of the **Algorithm** tab under **Settings**, click **Add Algorithm**.
2. Select **Payment Card**. The "Create Payment Card Algorithm" pane appears.
3. Enter an **Algorithm Name**.
Info: This MUST be unique.
4. Enter a **Description**.

5. Set **Minimum Masked Positions**. This value is the minimum number of positions that must be replaced for masking to be considered successful. If fewer positions are masked, a non-conforming data handling error is triggered. Values for this field must be in the range [0-32].
6. Set **Preserve Starting Digits**. This value specifies how many maskable characters should be preserved from the beginning of the input. Only maskable characters are included in this count. Values for this field must be in the range [0-32].
7. When you are finished, click **Save**.

For information on creating Payment Card algorithms through the API, see [API Calls for Creating Algorithms - Payment Card](#).

Examples

As an example, a Payment Card algorithm with a *minMaskedPositions* value of 6 and a *preserve* value of 6 may mask as follows:

- "5419033646326699" → "5419036803270758"
- "5419-0336-4632-6699" → "5419-0368-0327-0758"
- "5319abc0339def4632ghi6599!" → "5319abc0364def1507ghi4137!"

All inputs with the same sequence of digits masked with the same algorithm configuration will result in the same output values.

Regex Decompose (Algorithm frameworks)

Extensible Algorithm Framework

The Regex Decompose framework masks values that match specified [Java 8 regular expressions](#). The algorithm attempts to match the algorithm input against each regular expression, and once a match is found, the associated action is applied to transform either the entire input, or each capturing group (parts of the input) defined by the expression. A fallback action may be provided for use when none of the defined regular expressions match the input. If no fallback action is defined and an input fails to match any of the defined regular expressions, the algorithm may be configured to generate a non-conformant data exception.

Capturing groups are used in regular expressions to create subgroups. These can be expressed in regular expressions using parentheses to group characters together. This algorithm allows for different capturing groups to be assigned different mask actions. Nested capturing groups are unsupported and may lead to unpredictable behavior. If no capturing groups are defined, the first action is applied to the entire match. In this case, the action list should contain only one action.

Creation of Regex Decompose algorithms can only be done through the API, see [API Calls for Creating Algorithms - Regex Decompose](#).

Examples

As an example, a Regex Decompose algorithm with the following configuration:

```
Mask Pattern:
  Regular Expression: "[0-9]*)"
  Action: Redact
  Redact String: "redacted"
  Require Mask: false
  Trim Input: true
  Maximum Input Length: 10
```

Will produced masked results as follows:

- "12345" → "redacted"
- "6789" → " redacted "
- "12345678901" → non-conformant data
 - exceeds maximum input length
- "abc123" → "abc123"
 - remains unmasked since it does not match the regex pattern

The provided regular expression matches any inputs with 0 or more digits in the range [0-9] and any inputs that match will be replaced with the string "redacted". Any inputs that contain characters outside of the range [0-9] will not be masked. If require mask was set to true, the last example "abc123" would trigger a non-conformant data event as the value would not be masked by the algorithm.

Another example that includes capturing groups with the following configuration:

```
Mask Pattern:
  Regular Expression: "([1-9]*)-([a-z]*)"
  Action 1: Redact
  Redact Character: 'X'
  Action 2: Preserve
  Require Mask: true
  Trim Input: true
```

```
Maximum Input Length: 10
Fallback Action: Redact
Redact String: "redacted"
```

Will produce masked results as follows:

- "12345-abc" → "XXXXX-abc"
- "abc-123" → "redacted"
 - does not match the pattern so the fallback action is applied
- "1-a" → "X-a"
- "-" → "redacted"
 - does match the pattern but the masked output would be "-" which breaks the requirement that the output must be different from the input so the fallback action is applied
- "redacted" → non-conformant data
 - does not match the pattern so the fallback action is applied but the fallback action does not change the value so it fails the requirement that the input must be masked

The provided regular expression matches any inputs with 0 or more digits in the range [1-9], a dash, and 0 or more characters in the range [a-z]. Any inputs that do not match that pattern will be masked by the fallback action. If the fallback action fails to change the input, a non-conformant data event will occur.

All inputs with the same input value masked with the same algorithm configuration will result in the same output values.

Secure Lookup (Algorithm frameworks)

Extensible Algorithm Framework

Secure Lookup is the most commonly used type of algorithm. It is easy to generate and works with different languages. When this algorithm replaces real, sensitive data with fictional data, it is possible that it will create repeating data patterns, known as “collisions.” For example, the names “Tom” and “Peter” could both be masked as “Matt”. Because names and addresses naturally recur in real data, this mimics an actual data set. However, if you want the Masking Engine to mask all data into unique outputs, you should use Character Mapping.

Starting in version 6.0.4.0, we introduced a built in Extensible Secure Lookup Algorithm Framework. The new framework uses SHA256 hashing method and allows case configurations for input and output (i.e. masked) values.

Creating a secure lookup algorithm via UI

1. In the upper right-hand corner of the **Algorithm** tab, click **Add Algorithm**.
2. Choose **Secure Lookup Algorithm**. The Create SL Algorithm pane appears.
3. Enter an **Algorithm Name**.

Info

This MUST be unique.

4. Enter a **Description**.
5. Choose the **Output (Masked) Case** configuration. It is explained with the examples in the information popup window, which may be opened by clicking on the blue question sign on the above Create SL Algorithm window:

Secure Lookup

This is the framework to cover the scenarios where it is required to mask string using a lookup File.

Framework Options:

Output (Masked) Case:

- **Preserve Lookup File Case** - keep masked value as found in Lookup File
- **Preserve Input Case (Default)**- check the input case, which can be one of following three:
 - All uppercase - in that case force whole masked value to uppercase
 - All lowercase - in that case force whole masked value to lowercase
 - Mixed (if at least 1 character case is different from others) - in that case keep masked value as found in Lookup File
- **Force all Uppercase** - forces whole masked value to uppercase
- **Force all Lowercase** - forces whole masked value to lowercase

Hash Methods: A flag to configure hashing method for lookup determination. **default** is SHA256.

Case Sensitive Lookup: A flag to configure if input value case should be considered for Lookup match or not. **default** is off/false.

OK

6. Choose the **Hash Method** configuration.
 - **SHA256:** This hash method is the default hash method for extensible secure lookup algorithms.
 - **LEGACY:** This hash method is used to mimic the legacy secure lookup behavior in the extensibility framework.
7. Choose the **Case Sensitive Lookup** configuration. If **Case Sensitive Lookup** box is marked then the same input of different cases will be masked to the different values. For example:

```
Peter -> John
peter -> Andrew
```

If that setting is not marked (which is a default option), then lookup would be case insensitive, for example:

```
Peter -> John
peter -> John
```

8. Specify a **Lookup File**. This file is a single list of values that does not require a header, every line of the Lookup File might be used as a masked value. The Lookup File must be ASCII or UTF-8 encoding compatible. The lookup file can be referenced locally or with a specified/uploaded URI. The following is sample file content:

```
Smallville
Clarkville
Farmville
Townville
Cityname
Citytown
Towneaster
```

9. When you are finished, click **Save**.
10. Before you can use the algorithm in a profiling job, you must add it to a domain.

For information on creating Secure Lookup algorithms through the API, see [API Calls for Creating Algorithms - Secure Lookup](#).

Segment Mapping (Algorithm frameworks)

Extensible Algorithm Framework

Segment Mapping algorithms produce no overlaps or repetitions in the masked data. They let you create unique masked values by dividing a target value into separate segments and masking each segment individually.

You might use this method if you need columns with unique values, such as Social Security Numbers, primary key columns, or foreign key columns. When using segment mapping algorithms for primary and foreign keys, in order to make sure they match, you must use the same Segment Mapping algorithm for each. You can set the algorithm to produce alphanumeric results (letters and numbers) or only numbers.

With Segment Mapping, you can set the algorithm to ignore specific characters. For example, you can choose to ignore dashes [-] so that the same Social Security Number will be identified no matter how it is formatted. You can also preserve certain values. For example, to increase the randomness of masked values, you can preserve a single number such as 5 wherever it occurs. Or if you want to leave some information unmasked, such as the last four digits of Social Security numbers, you can preserve that information.

This algorithm can be used for tokenization and re-identification jobs if the following conditions are met:

- All alpha-numeric and numeric segments have Value Ranges with "Mask values with: The same ranges"
- There are no segments with "Segment Treatment: Mask with a constant value"
- If a numeric segment is defined, "Short Numeric Segment Handling: Report nonconforming data" is selected

To decide whether Character Mapping or Segment Mapping is the correct option for your use case, see [Choosing Between Character and Segment Mapping Frameworks](#).

Creating a segment mapping algorithm via UI

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Numeric Expression
- Extended

Create Segment Mapping Algorithm [Learn More](#)

Algorithm Name

Description

Segment 1

+

Short Numeric Segment Handling

Report nonconforming data ▼

Ignore Characters

Automatically ignore special characters ▼

Process Preserve Segments Before Ignore Characters

Cancel

Save

1. In the upper right-hand region of the **Algorithms** tab, click **Add Algorithm**.
2. Select **Segment Mapping**. The "Create Segment Mapping Algorithm" pane appears.
3. Enter an **Algorithm Name**.
Info
This MUST be unique.
4. Enter a **Description** (optional).
5. Click the **Segment 1** tab to open the pane for the first segment. Use the plus (+) button to add as many segments as you need (maximum of 10). Use the tabs to navigate between segments.

Segment 1 +

Length: 1 Segment Treatment: Mask alpha-numeric

Value Ranges ?

Alpha-Numeric Ranges and Values
 If original values are: A-E,X,W-Z,1-5,7
 Mask values with: Different ranges
Alpha-Numeric Ranges and Values
 Replace values with: A-E,X,W-Z,1-5,7

6. For each segment, select its:
- **Length** (number of characters). The maximum is 6.
 - **Segment Treatment:** Mask alpha-numeric, Mask numeric, Preserve, or Mask with a constant value.
 - **Value Ranges.** Optional for alpha-numeric and numeric, required for constant. See [Specifying Value Ranges](#).
- Info**
Numeric segments are masked as whole segments. **Alpha-numeric** segments are masked by individual characters.
7. If you would like to allow the masking of short numeric segments, change the **Short Numeric Segment Handling** drop-down to select **Mask partial segments**. This option allows masking to proceed if an input string is truncated midsegment. For example, you define a numeric segment of length 4, but the input string ends midsegment so you have a 2 digit number instead of 4.
- Info**
 This only applies to **Mask numeric** segments. Other segment treatments always apply to partial segments.
- Info**
 If **Mask partial segments** is selected AND a **Mask numeric** segment is defined, the algorithm is not reversible and cannot be used for tokenization/re-identification.
 By default, the segment mapping algorithm will **Report nonconforming data** for short numeric segments and the **Monitor** page will display a warning that can be used to report the non-conformant data events. This will result in the non-conformant data not being masked.
- Example

Your content goes here

Segment 1: length 2, mask alpha-numeric. **Segment 2:** length 4, mask numeric.

Input	Output	Short Numeric Segment Handling
AB1234	DL9148	Either
AB12	AB12	Report nonconforming data (reported)
AB0012	DL3619	Report nonconforming data (not reported)
AB12	DL3619	Mask partial segments

- Select the appropriate **Ignore Characters** handling. Ignored characters are removed from the input value before masking and restored to their original positions after masking. When **Automatically ignore special characters** is selected, all non-maskable characters are ignored. When **Ignore specific characters** is selected, only specified characters are ignored. Enter the characters you wish to ignore in the **Specific Characters** box, separated by a comma. To ignore the comma character (,), check the **Ignore Commas** checkbox. To ignore control characters, check the **Add Control Characters** checkbox and select the desired characters to ignore.

Ignore Characters

Ignore specific characters
▼

Specific Characters (separated by comma)

@,#,-,+

Ignore Commas

Add Control Characters

- | | | | |
|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| <input type="checkbox"/> ^@ [NUL] | <input type="checkbox"/> ^H [BS] | <input type="checkbox"/> ^P [DLE] | <input type="checkbox"/> ^X [SUB] |
| <input type="checkbox"/> ^A [SOH] | <input type="checkbox"/> ^I [TAB] | <input type="checkbox"/> ^Q [DC1] | <input type="checkbox"/> ^Y [ESC] |
| <input type="checkbox"/> ^B [STX] | <input type="checkbox"/> ^J [LF] | <input type="checkbox"/> ^R [DC2] | <input type="checkbox"/> ^Z [CAN] |
| <input type="checkbox"/> ^C [ETX] | <input type="checkbox"/> ^K [VT] | <input type="checkbox"/> ^S [DC3] | <input type="checkbox"/> ^_ [GS] |
| <input type="checkbox"/> ^D [EOT] | <input type="checkbox"/> ^L [FF] | <input type="checkbox"/> ^T [DC4] | <input type="checkbox"/> ^^ [RS] |
| <input type="checkbox"/> ^E [ENQ] | <input type="checkbox"/> ^M [CR] | <input type="checkbox"/> ^U [NAK] | <input type="checkbox"/> ^[[EM] |
| <input type="checkbox"/> ^F [ACK] | <input type="checkbox"/> ^N [SO] | <input type="checkbox"/> ^V [SYN] | <input type="checkbox"/> ^\ [US] |
| <input type="checkbox"/> ^G [BEL] | <input type="checkbox"/> ^O [SI] | <input type="checkbox"/> ^W [ETB] | <input type="checkbox"/> ^/ [FS] |

- Lastly, the checkbox for **Process Preserve Segments Before Ignore Characters** selects whether to process segments with "Segment Treatment: Preserve" first, before removing ignore characters, so ignore characters count as length when finding preserve segments in the input, and then process the remaining

segments.

The default is for this to be unchecked, so ignore characters are removed first, and then the segments are processed in order.

Warning:

This option exists to support backwards compatibility with the legacy Segment Mapping algorithm configuration and is **not recommended** for newly created algorithms, as it may cause some segments to be processed out of order.

10. When you are finished, click **Save**.
11. Before you can use the algorithm in a profiling job, you must add it to a domain. If you are not using the Masking Engine Profiler to create your inventory, you do not need to associate the algorithm with a domain.

Specifying value ranges

You can specify values ranges for each segment based on the **Segment Treatment**.

For **Mask alpha-numeric**, you can specify an original value range and a mask value range. If either of these fields is left blank, it will use the default value range, which is 0-9,A-Z. Use the value range fields to specify individual values and ranges, for example 'A-F,P,R,1-5,7,9'.

 The masking will only look to mask these values and will preserve any other values. Letters are masked to letters and digits to digits.

 If the original and replacement values and ranges are not the same, the algorithm is not reversible and cannot be used for tokenization/re-identification.

For **Mask numeric**, you can specify an original value range and a mask value range. If either of these fields is left blank it will use the default value range, which is 0 to the max integer that can fit into the segment length (ex: 000-999 for a segment of length 3). Use the value range fields to specify integer values and ranges, for example '10,30,50-875'.

 The masking will only look to mask these values and will preserve any other values.

For **Preserve**, you cannot specify any value ranges as whatever is encountered in this segment will be preserved.

For **Mask with a constant value**, you cannot specify an original value range, and your replace value must be a single value the same length as the segment (ex: if the segment length is 3, 'ABC' would be valid replacement).

 The Segment Mapping pattern and sub-patterns need to match the data in order for it to be masked. If the data is longer than the defined pattern it will be passed through unmasked. To avoid this unwanted behavior - patterns (segments) and Ignore Characters should be set to match the data.

For information on creating Segment Mapping algorithms through the API, see [API Calls for Creating Algorithms - Segment Mapping](#).

Examples

Perhaps you have an account number for which you need to create a segment mapping algorithm. You can separate the account number into segments, preserving the first two-character segment, replacing a segment with a specific value, and preserving a hyphen. The following is a sample value for this account number:

NM831026-04

Where:

- **NM** is a plan code number that you want to preserve, always a two-character alphanumeric code.

- **831026** is the uniquely identifiable account number. To ensure that you do not inadvertently create actual account numbers, you can replace the first two digits with a sequence that never appears in your account numbers in that location. (For example, you can replace the first two digits with 98 because 98 is never used as the first two digits of an account number.) To do that, you want to split these six digits into two segments. The first of these segments would be a 2 character constant segment mapping to 98. The second of these 2 could be a 4 character numeric segment.
- **-04** is a location code. You want to preserve the hyphen and you can replace the two digits with a number within a range (in this case, a range of 1 to 77).

Tokenization (Algorithm frameworks)

Extensible Algorithm Framework

The Tokenization framework allows you to mask data and reverse its masking. For example, you can use a Tokenization algorithm to mask data before you send it to an external vendor for analysis. The vendor can then identify accounts that need attention without having any access to the original, sensitive data. Once you have the vendor's feedback, you can reverse the masking and take action on the appropriate accounts.

The Tokenization algorithm is designed to be used in Tokenization/Re-Identification jobs, though it can also be used in Masking.

The algorithm tokenizes values using AES-128 encryption in CBC-CTS mode, with an optional initialization vector (IV), and Base64 encoding. The results are alpha-numeric strings that are longer than the original values. If the result is too long to fit in the field, the algorithm can be configured to either (a) fallback to a reversible masking algorithm, which produces a result that is the same length as the original value, or (b) fail the job.

The algorithm has the following properties:

- The masked value for each input is consistent when using the same algorithm **and** the initialization vector length is 0. Changing the key for the algorithm or using an initialization vector length greater than 0 will result in different masked values.
- As long as at least one maskable character is present in the input, the masked value will never match the input.
- The algorithm used to mask a value can change depending on the length of the input.
- The algorithm only works on string data types. Numbers can be masked if the column data type is a String type, such as VARCHAR or TEXT.

This new algorithm framework was introduced in version 6.0.13.0 to replace the existing Tokenization algorithm and adds the ability to select a fallback algorithm.

Creating a tokenization algorithm via UI

1. In the upper right-hand region of the **Algorithms** tab under **Settings**, click **Add Algorithm**.
2. Select the **Tokenization** Framework. The "Create Tokenization Algorithm" pane appears.

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping (legacy)
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Extended

Create Tokenization Algorithm

Algorithm Name

Description

Initialization vector length

16 (default) ▼

Fallback

None ▼

Cancel

Save

1. Enter an **Algorithm Name**.
Info: This MUST be unique.
2. Enter a **Description**.
3. Select an **Initialization vector length**. The default length is 16, which offers the most security. The tradeoff is that this increases the length of the masked result. Selecting a lower IV length decreases the length of the masked result. It is recommended that you only select an IV length of 0 if you require the masked value for each input to be consistent between jobs and for the same input to only mask to one output.
4. Select a **Fallback** algorithm. An AES encrypted result is always longer than the original value. If an AES encrypted result is too long to fit into the field, the job will fail if Fallback is "None". When Fallback is "Character Mapping", the Character Mapping algorithm is used to tokenize the value, which produces a result that is the same length as the input.

If **Character Mapping** is selected as the Fallback, a Character Mapping algorithm is created, which will be used to tokenize values that cannot be tokenized with AES encryption because the encrypted result is too long for the field. When selected, two additional configuration options will appear: **Minimum Masked Positions** and **Character Groups**. Unlike standalone Character Mapping algorithms, the Character Mapping algorithm used for Tokenization fallback does not support **Preserve Ranges** and **Preserve Leading Zeroes**, and **Case Sensitive** is permanently set to **true**.

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping (legacy)
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Extended

Create Tokenization Algorithm

Algorithm Name

Description

Initialization vector length

Fallback

Minimum Masked Positions

Character Groups [? Learn More](#)

✕

1. Enter a value for **Minimum Masked Positions**, which sets the minimum number of characters that the algorithm must mask; fewer positions triggers non-conformant data handling. Null, empty, and all-whitespace values never trigger non-conformant data handling.
2. Define **Character Groups** for each group of characters among which you would like to map. Each group may be defined either by specifying each literal character in the group, such as "0123456789", or using Java Regular Expression style character ranges, such as "[0-9]". The algorithm will freely map characters to other characters within the same group, so by defining groups "[0-9]" and "[A-Z]", numbers would be replaced by other numbers, and letters by other letters, but a number would never be replaced by a letter. Groups should not contain duplicate characters, and each character may belong to only one group. Any character that is not assigned to a group will be preserved (not masked) by the algorithm. It is recommended that all characters are in one group so there is more randomization and the values are more obfuscated. The default is the Base64 character set ["[A-Za-z0-9+/]"], which contains the same characters that appear in an AES encrypted result.

Once you have created an algorithm, you may associate it with a domain.

1. In the upper right-hand region of the **Domains** tab under **Settings**, click **Add Domain**.

Add Domain

Domain Name

Algorithm Name

Tokenization Algorithm Name

2. Enter a **Domain Name**.
3. Select algorithms from both the **Algorithm Name** and **Tokenization Algorithm Name** drop-down menus.

Next, create a Tokenization Environment:

1. In **Environments**, use the **Select Action** dropdown menu to select **Add Environment**.

Add Environment

Application Name

test

Environment Name

Tokenization ReIdentify QA

Purpose

Tokenize/Re-Identify

Enable Approval Workflow

Cancel

Save

Save & View

2. For **Purpose**, select **Tokenize/Re-Identify**. **Info** This environment will also be used to re-identify your data.
3. Set up a Tokenization job using the Tokenization Method. Execute the job.

Create Tokenization Job

<p>Job Name</p> <input style="width: 90%;" type="text" value="QA Tokenize"/>	<p>Commit Size</p> <input style="width: 90%;" type="text"/>	<p>Feedback Size</p> <input style="width: 90%;" type="text"/>				
<p>Tokenization Method</p> <input style="width: 90%;" type="text" value="Tokenization Method"/>	<p><input type="checkbox"/> Disable Trigger</p> <p><input type="checkbox"/> Batch Update <input type="checkbox"/> Disable Constraint</p> <p><input type="checkbox"/> Drop Indexes</p>					
<p>Target: Tokenization Re...</p> <p><input type="checkbox"/> Multi Tenant</p> <p>Rule Set</p> <input style="width: 90%;" type="text" value="Rule Set"/>	<p>Prescript</p> <input style="width: 90%;" type="text" value="Select..."/> <p>Postscript</p> <input style="width: 90%;" type="text" value="Select..."/>					
<p>Streams:</p> <table border="0" style="width: 100%;"> <tr> <td style="width: 50%;">Number</td> <td style="width: 50%;">Row Limit</td> </tr> <tr> <td><input style="width: 90%;" type="text" value="20"/></td> <td><input style="width: 90%;" type="text"/></td> </tr> </table>	Number	Row Limit	<input style="width: 90%;" type="text" value="20"/>	<input style="width: 90%;" type="text"/>	<p>Comments</p> <div style="border: 1px solid #ccc; height: 40px; width: 100%;"></div>	
Number	Row Limit					
<input style="width: 90%;" type="text" value="20"/>	<input style="width: 90%;" type="text"/>					
<table border="0" style="width: 100%;"> <tr> <td style="width: 50%;">Min Memory</td> <td style="width: 50%;">Max Memory</td> </tr> <tr> <td><input style="width: 90%;" type="text" value="In MB"/></td> <td><input style="width: 90%;" type="text" value="In MB"/></td> </tr> </table>	Min Memory	Max Memory	<input style="width: 90%;" type="text" value="In MB"/>	<input style="width: 90%;" type="text" value="In MB"/>	<p>Email</p> <div style="border: 1px solid #ccc; height: 40px; width: 100%; text-align: right; padding-right: 10px;"> </div>	
Min Memory	Max Memory					
<input style="width: 90%;" type="text" value="In MB"/>	<input style="width: 90%;" type="text" value="In MB"/>					
<p>Update Threads</p> <input style="width: 90%;" type="text" value="4"/>						
<p>If Nonconforming Data is encountered</p> <p><input type="checkbox"/> Stop job on first occurrence</p>						
<input style="width: 100px;" type="button" value="Cancel"/>		<input style="width: 100px;" type="button" value="Save"/>				

Examples

Here is example data before and after Tokenization:

Before Tokenization

```

1,Erasmus,245 Park Ave,123-45-6789
2,Salathiel,245 park ave,123-45-6789
3,Salathiel,1003 Stant Drive,111-11-1111
    
```

After Tokenization

ID, fname, address, ssn

1, FQL71CmqK/pkd8B2vVP90304+/
krT91dscS0rKQRACQ=, XFLst0IcSb0a2Ule0mlACPkca0EVczZsEdxl225kF1M=, x6tJ4eyL4it4ji84h8Pzo
CW4QBZphEqD0y3hEj4h1jE=

2
, 4bGZoCLpbV2zAMsTkcc5lMTBKksvOP+tfAWucq+BnKM=, 0A9dJ5HN5oRx18ZY01f5Y8DofvhFoRo98cuQH7
YeEo=, Evj+LnETt7ABbXlTDPyNvvJe8WJnrhEWeS0lqtqrr4U=

3

, L14T49FrCBYRib0AK0Y4vbnsbw0n1RpqBU97EGg4RvA=, f6AR0T+HBoTW7+l0e8ok9rImj872PUnYYNYMDYS
y4dw=, wYmvEhktV371kqH607afJHZloT+4DYNJxehWicPZJzI=

General UI for extended algorithms

Overview

An algorithm plugin can be configured through the graphical user interface by entering the plugin's required configuration in JSON format. The following section describes how to use this feature.

GUI steps

- Use the **Select Framework** drop-down to create an instance corresponding with the selection.
- Provide an **Algorithm Name** (Required).
- Provide a **Description** for the new algorithm instance (Optional).
- Provide a valid extension of the corresponding framework in JSON format in **Configuration JSON** (Required).
- Based on the Select Framework option, the Configuration JSON will be populated with default values in the corresponding text area.
- A **Help** icon will appear to show the selected framework details and configuration schema.
- A **Utility** icon on top of the configuration JSON text area is available to upload and copy the local file reference, and to pick an algorithm reference from existing algorithm instances.
- The **Format JSON** button is used to format the text from the Configuration JSON text area into JSON format.
- The **Validate Configuration** button will validate the Configuration JSON format, and also validate against the selected framework configuration schema.
- For a plugin with a specific GUI like Character Mapping or Secure Lookup, their respective GUI will be shown when editing.
- For other plugin instances, the user can only modify the description and extension of the algorithm instance from the plugin GUI. The select framework and algorithm name fields will be read-only.

The default selected framework populates corresponding Configuration JSON in the text area.

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping (legacy)
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Extended**

Create Extended Algorithm

Select Framework [details](#)

Character Mapping

Algorithm Name

Description

Configuration JSON

```
{  
  "characterGroups": null,  
  "caseSensitive": false,  
  "minMaskedPositions": 1,  
  "preserveRanges": null,  
  "preserveLeadingZeros": false  
}
```

Format JSON Validate Configuration Cancel Save

When the framework changes, the Configuration JSON will be populated automatically.

Select Framework

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping (legacy)
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Extended**

Create Extended Algorithm

Select Framework [details](#)

FullName

Algorithm Name

Description

Configuration JSON ⚙️

```
"lastNameSeparators": [
  ""
],
"lastNameAtTheEnd": true,
"maxLengthOfMaskedName": 0,
"maxNumberFirstNames": 2,
"firstNameAlgorithmRef": {
  "name": "dlpx-core:FirstName"
},
"lastNameAlgorithmRef": {
  "name": "dlpx-core:LastName"
}
}
```

utility

Format JSON Validate Configuration Cancel Save

Framework details will appear.

FullName Framework

Plugin Details:

pluginId: 7

pluginName: dlpx-core

pluginAuthor: Delphix Engineering

pluginType: EXTENDED_ALGORITHM

Configuration Schema:

```
{
  "id": "urn:jsonschema:algorithm:plugin:name:FullName",
  "properties": {
    "ifSingleWordConsiderAsLastName": {
      "type": "boolean",
      "description": "If single word - consider as last name (default) or first name."
    },
    "lastNameSeparators": {
      "type": "array",
      "description": "A list of characters that serve as last name separator.\nEverything prior to the first detected separator - is a last name.",
      "items": {
        "type": "string"
      }
    },
    "lastNameAtTheEnd": {
      "type": "boolean",
      "description": "If last name separator is not detected - this property defines"
    }
  }
}
```

If there are issues with the Configuration JSON, an **Invalid input** banner will appear.

Invalid input network **Create Extended Algorithm**
JSON.parse: end of data after property value in object at line 15 column 1 of the JSON data

Secure Lookup

Character Mapping

Payment Card

Date

Dependent Date Shift

Name

Full Name

Email

Segment Mapping (legacy)

Mapping

Binary Lookup

Tokenization

Min Max

Data Cleansing

Free Text Redaction

Extended

Select Framework [details](#)

FullName

Algorithm Name

Description

Configuration JSON

```
"lastNameSeparators": {  
  "."  
},  
"lastNameAtTheEnd": true,  
"maxLengthOfMaskedName": 0,  
"maxNumberFirstNames": 2,  
"firstNameAlgorithmRef": {  
  "name": "dlpx-core-FirstName"  
},  
"lastNameAlgorithmRef": {  
  "name": "dlpx-core-LastName"  
}
```

Format JSON Validate Configuration Cancel Save

If the Configuration JSON is valid, a **Success** banner will appear.

Success framework Create Extended Algorithm

valid Extension

- Secure Lookup
- Character Mapping
- Payment Card
- Date
- Dependent Date Shift
- Name
- Full Name
- Email
- Segment Mapping (legacy)
- Mapping
- Binary Lookup
- Tokenization
- Min Max
- Data Cleansing
- Free Text Redaction
- Extended

Select Framework [details](#)

FullName

Algorithm Name

Description

Configuration JSON ⚙️

```

{
  "lastNameSeparators": [
    "."
  ],
  "lastNameAtTheEnd": true,
  "maxLengthOfMaskedName": 0,
  "maxNumberFirstNames": 2,
  "firstNameAlgorithmRef": {
    "name": "dipx-core:FirstName"
  },
  "lastNameAlgorithmRef": {
    "name": "dipx-core:LastName"
  }
}
            
```

The Plugin Helper Utility offers a way to upload a file and receive a reference id for algorithm extension, or to select an instance of algorithm instances for algorithm chaining.

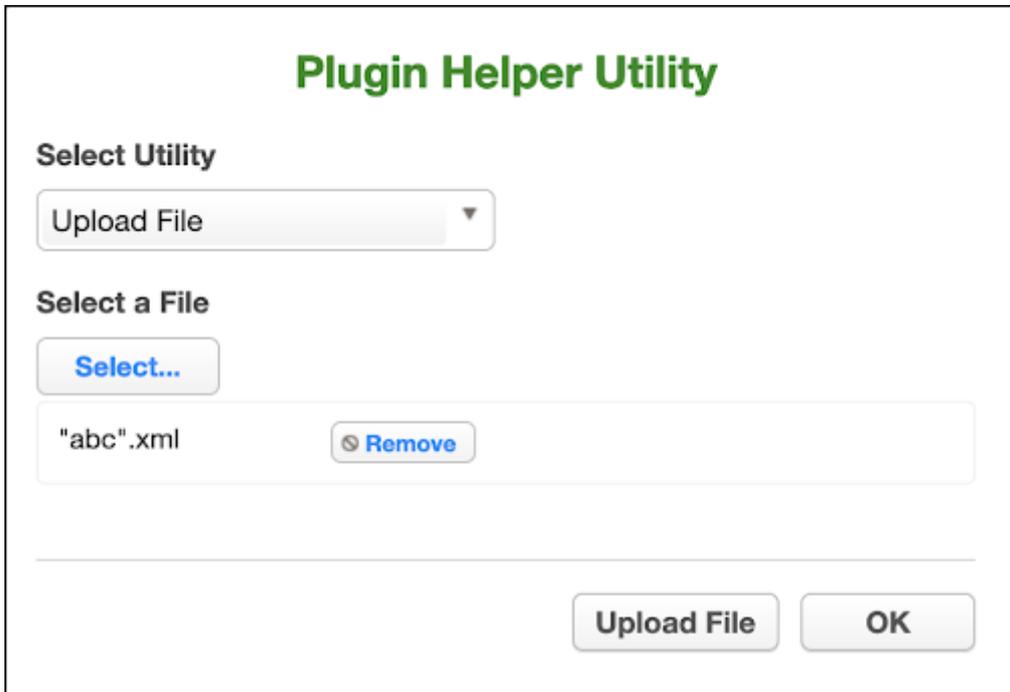
Plugin Helper Utility

Select Utility

Upload File

Select a File

In the Plugin Helper Utility, choose **Upload File** from the Select Utility drop-down. Select a file to upload, then click the **Upload File** button.



The screenshot displays the 'Plugin Helper Utility' interface. At the top, the title 'Plugin Helper Utility' is shown in green. Below the title, there is a section labeled 'Select Utility' with a dropdown menu currently set to 'Upload File'. Underneath, the 'Select a File' section contains a 'Select...' button. A file named '"abc".xml' is listed in a box, with a 'Remove' button next to it. At the bottom right of the utility, there are two buttons: 'Upload File' and 'OK'.

When the file is uploaded, it will render a copyable Value.

Plugin Helper Utility

Select Utility

Upload File ▾

Select a File

Select...

Test copy 3.xml Remove

Value

```
{ "uri": "delphix-file://upload/f_619a5327bcf3495ba43212714a5e69e1/Test%20copy%203.xml" }
```

copy

Upload File OK

The Plugin Helper Utility also has a **Select Algorithm** option in the Select Utility drop-down, which renders a new list of available algorithms to select.

Plugin Helper Utility

Select Utility

Select Algorithm ▼

Select Algorithm

dlpx-core:CM Numeric ▼

Value

dlpx-core:CM Numeric 

copy

OK

Builtin Driver Supports

Introduction

In 6.0.11.0, Delphix introduced the first built-in driver support plugin for the Oracle database platform.

The native connector types with a built-in driver support plugin are:

Native Connector Type	Release
Oracle	6.0.11.0
MSSQL	6.0.12.0
PostgreSQL	11.0.0.0

The built-in driver support plugins replace and improve upon the native connector database masking options of Disable Constraints, Drop Indexes, and Disable Triggers that have long had issues with functionality and negatively affecting job performance. Delphix has implemented the built-in driver support plugin for native connectors with Disable Constraints, Drop Indexes, and Disable Triggers tasks using the [Driver support plugin framework](#) released in 6.0.9.0. These optimizations apply to masking, reidentification, and tokenization jobs where these tasks are enabled.

For details on how to enable/disable these tasks on supported native connector jobs using the new Driver Support Plugin Framework, see [API Calls for managing masking job driver support tasks](#).

To retrieve information about job failures due to driver support task failures, an execution event will be raised and is accessible via the `GET /execution-events` endpoint: 1. `eventType` -

`DRIVER_SUPPORT_TASK_FAILURE` 2. `exceptionDetail` - Error message about the task failure that will typically include the error code that is specific to the database platform

Oracle

For details on usage and known limitations of the Oracle Disable Constraints, Drop Indexes, and Disable Triggers driver support tasks, see [Oracle Built-in driver support plugin](#).

MSSQL

For details on usage and known limitations of the MSSQL Disable Constraints, Drop Indexes, and Disable Triggers driver support tasks, see [MSSQL Built-in driver support plugin](#).

PostgreSQL

For details on usage and known limitations of the PostgreSQL Drop Constraints, Drop Indexes, and Disable Triggers driver support tasks, see [PostgreSQL Built-in Driver Support Plugin](#).

Built-in Oracle driver support plugin

For instructions on how to enable/disable Disable Constraints, Drop Indexes and Disable Triggers on Oracle jobs, see [API Calls for managing masking job driver support tasks](#).

Optimizations

 Disable Constraints disables and re-enables constraints *while keeping the index associated with the constraint*. In order to drop and re-create the index associated with the constraint, enable Drop Indexes along with Disable Constraints.

For **in-place** jobs:

1. **Disable Constraints** disables and re-enables constraints on only masked columns.
2. **Drop Indexes** drops and re-creates indexes on only masked columns.
3. **Disable Triggers** disables and re-enables triggers on only tables with masked columns.

For **on-the-fly** jobs, the tasks will execute on all columns/tables in the ruleset.

Task execution order

 The order of task execution for built-in driver support plugins is fixed/unmodifiable.

The order of the tasks is as follows:

Pre-job:

1. Disable Constraints
2. Drop Indexes
3. Disable Triggers

Post-job (mirrored order):

1. Disable Triggers
2. Drop Indexes
3. Disable Constraints

Enabling tasks on a job

For instructions on how to enable driver support tasks on jobs, see [API calls for managing masking job driver support tasks](#).

Important considerations

1. If masking primary key fields:
 - a. Use the same deterministic algorithms on primary key fields that reference each other, so that referential integrity is maintained when the masking transformation completes and all constraints are re-enabled.
 - b. Enable both Disable Constraints and Drop Indexes.
2. If dropping indexes on masked fields with constraints is desired, enable both Disable Constraints and Drop Indexes. The implementation of the optimizations has been modified, such that Disable Constraints only disables constraints and keeps indexes automatically created and Drop Indexes handles dropping/recreating indexes. The change in task order and separation of concerns with the functionality of the tasks

resolves issues around missing indexes present with the legacy database masking option of Disable Constraints.

Known limitations

1. If masking a primary key field, if only Disable Constraints is enabled, the job will fail during the transformation. It is recommended to enable both Disable Constraints and Drop Indexes on any applicable job per the usage instructions above. In order to not have Drop Indexes enabled, adding a prescript that disables the desired constraints will also work, but note that this workaround may result in missing indexes.
2. Delphix supports the below indexes:
 - Normal indexes
 - Functional indexes (6.0.12.0 and later)
 - Local, global, and partial partition indexes (6.0.15.0 and later)

Built-in MSSQL driver support plugin

Summary

The current implementation is simply the earlier implementation of the database masking options in the new driver support plugin framework. No optimizations have been implemented yet; stay tuned for optimizations in a future release.

Tasks

For **in-place** jobs:

1. **Disable Constraints** disables and re-enables constraints on all columns of the table(s) included in the job ruleset.
2. **Drop Indexes** drops and re-creates indexes on only masked columns.
3. **Disable Triggers** disables and re-enables triggers on all tables included in the job ruleset.

For **on-the-fly** jobs, the tasks will execute on all columns in all tables included in the ruleset.

Task execution order

 The order of task execution for built-in driver support plugins is fixed/unmodifiable.

The order of the MSSQL Driver Support tasks is as follows:

preJob:

1. Disable Constraints
2. Drop Indexes
3. Disable Triggers

postJob:

1. Disable Triggers
2. Drop Indexes
3. Disable Constraints

Enabling tasks on a job

For instructions on how to enable driver support tasks on jobs, see [API calls for managing masking job driver support tasks](#).

Known limitations

1. Primary Key constraints are not disabled.
2. Unique Constraints/Indexes are not disabled.
3. Clustered Column store Indexes are not dropped.
4. Functional Indexes are not dropped.
5. As before, constraints are dropped on all columns of the table(s) included in the job ruleset.
6. Referential integrity is not enforced, i.e., in the current implementation, there is no validation that a primary key or unique constraint column being referenced by a foreign key column are masked with the same deterministic algorithm.

7. Disable Triggers is dropping the triggers on all tables in the ruleset irrespective of whether table is masked or not.

Built-in PostgreSQL driver support plugin

For instructions on how to enable/disable Drop Constraints, Drop Indexes, and Disable Triggers on PostgreSQL jobs, see [API Calls for Managing Masking Job Driver Support Tasks](#).



Unlike other database platforms such as Oracle and MSSQL, PostgreSQL doesn't support the disabling of constraints. Therefore, this plugin has a Drop Constraints task instead of a Disable Constraints task.

Tasks

For **in-place** jobs:

1. **Drop Constraints** drops and re-creates constraints on masked columns only.
2. **Drop Indexes** drops and re-creates indexes on masked columns only (except for indexes automatically generated by PostgreSQL to support a primary key or unique constraint - those are dropped by Drop Constraints).
3. **Disable Triggers** disables and re-enables triggers on tables with masked columns only.

For **on-the-fly** jobs, the tasks will execute on all columns and tables in the ruleset.

Task Execution Order



The order of task execution for built-in driver support plugins is fixed/unmodifiable.

The order of the tasks is as follows:

Pre-job:

1. Drop Constraints
2. Drop Indexes
3. Disable Triggers

Post-job (mirrored order):

1. Disable Triggers
2. Drop Indexes
3. Drop Constraints

Important Considerations

1. If masking primary key fields, use the same deterministic algorithms on primary key fields that reference each other so that referential integrity is maintained when the masking transformation completes and all constraints are re-created.
2. If dropping indexes on masked fields with constraints is desired, enable both Drop Constraints and Drop Indexes. PostgreSQL automatically creates unique indexes to support primary key and unique constraints. Those indexes cannot be dropped by the Drop Indexes task, but enabling Drop Constraints will allow it to drop those indexes when the constraints they support are dropped.

Known Limitations

1. If both a foreign key column and the primary key or unique column it references are not masked with the same deterministic algorithm (guaranteeing referential integrity), the job will fail due to an integrity constraint violation.

-
2. If masking a field that is a primary key or has a unique constraint, and Drop Indexes is the only task enabled, the job will fail. Drop Constraints must also be enabled so that it can drop the indexes PostgreSQL automatically creates to support constraints.

Creating masking jobs

This section describes how users can create a masking job.

Creating new jobs

In the **Environment Overview** screen, select one of the jobs icons to create the corresponding job:

- Profile
- Mask

The screenshot displays the DELPHIX MASKING application interface. At the top, there is a blue navigation bar with the text 'DELPHIX MASKING' on the left, 'Job Wizard' and 'admin' on the right, and a menu with 'Environments', 'Monitor', 'Settings', 'Admin', and 'Audit'. Below this, there are tabs for 'Overview', 'Connector', 'Rule Set', and 'Inventory'. The 'Overview' tab is active, showing a breadcrumb 'Home > Environments > test1' and the title 'test1'. On the right side of the overview, there are three buttons: 'Export', 'Profile', and 'Mask'. Below these buttons is a section titled 'Environment' with a table of details:

Name	test1
Purpose	Mask
Application Name	test1
Approval workflow	Disabled

Below the environment details is a table of jobs:

Job ID	Name	Rule Set	Completed	Status	Action	Edit	Delete
1	ProfileJob	fileconnector	2021-10-04 09:32	★ Succeeded			

At the bottom of the page, there is a navigation bar with links for 'Environments', 'Monitor', 'Settings', 'Admin', and 'Audit', and the DELPHIX logo on the right.

Creating a new masking job

To create a new masking job:

1. Click **Mask**. The **Create Masking Job** window appears.

Create Masking Job

Job Name	<input type="text"/>	Commit Size	<input type="text"/>	Feedback Size	<input type="text"/>
Masking Method	<input type="text" value="Masking Method"/>	<input type="checkbox"/> Disable Trigger			
Target: test1		<input checked="" type="checkbox"/> Batch Update	<input type="checkbox"/> Disable Constraint		
<input type="checkbox"/> Multi Tenant		<input type="checkbox"/> Drop Indexes			
Rule Set	<input type="text" value="Rule Set"/>	Prescript	<input type="text" value="Select..."/>		
Streams:		Postscript	<input type="text" value="Select..."/>		
Number	<input type="text" value="1"/>	Row Limit	<input type="text"/>		
Min Memory	<input type="text" value="In MB"/>	Max Memory	<input type="text" value="In MB"/>		
Update Threads	<input type="text" value="1"/>	Comments	<input type="text"/>		
If Nonconforming Data is encountered	<input type="checkbox"/> Stop job on first occurrence				
		Email	<input type="text" value="manisha.gupta@delphix.com"/>		
		Cancel	Save		

2. You will be prompted for the following information:
 - a. **Job Name** – A free-form name for the job you are creating. Must be unique across the entire application.
 - b. **Masking Method** – Select either **In-Place** or **On-The-Fly**. **In-Place** jobs update the source environment with the masked values. **On-The-Fly** jobs read unmasked data from the source environment and writes the masked data to the target environment.

Info:
On-The-Fly Masking Jobs. Only certain combinations of connector types are supported. On-The-Fly jobs where the source and target connectors are of the same type (e.g. Oracle to Oracle, delimited file to delimited file), and jobs with a database source (e.g. Oracle, MS SQL) and the target is delimited files are supported. The target tables or files must be created in advance and the names must match the names of the source tables or files. In the case of a database to delimited file job, the file names should match the table names.
 - c. **Multi Tenant** – Checkbox if the job is for a multi-tenant database.

Info: Provisioning Masked VDBs.

A job must be Multi-Tenant to use it when creating a masked virtual database (VDB). This option allows existing rulesets to be reused to mask identical schemas via different connectors. The connector can be selected at job execution time.

- d. **Rule Set** — Select a rule set that this job will execute against.
- e. **Source Environment** (only for On-The-Fly Masking Method) - Select the Source Environment that this job will get the data from.
- f. **Source Connector** (only for On-The-Fly Masking Method) - Select the Source Connector that provides the connection to the the chosen Source Environment.
- g. **Streams: Number**—The number of parallel streams to use when running the job. For example, you can select two streams to mask two tables in the Rule Set concurrently in the job instead of one table at a time.

Info:**Choosing the number of streams**

Jobs - even with a single stream - will have separate execution threads for input, masking, and output logic. While it is not necessary to increase the number of streams to engage multiple CPU cores in a job, doing so may increase overall job performance dramatically, depending on a number of factors. These factors include the performance characteristics of the data source and target, the number of processor cores available to the Delphix Masking Engine, and the number and types of masking algorithms applied in the Rule Set. The memory requirements for a job increase proportionately with the number of streams.

- h. **Streams: Row Limit**—The number of data rows that may be in process simultaneously for each masking stream. For file jobs, this controls the number of delimited or fixed-width lines, mainframe records, or XML elements in process at one time. Setting this value to 0 allows unlimited rows into each stream, while leaving it blank will select a default limit based on job type.

Info:**Choosing the Row Limit**

The default Row Limit values have been selected to allow typical jobs to run successfully with the default job memory and streams number settings. This assumes a maximum row or record size of approximately 2000 bytes with 100 masked columns. If masked row or record size, or column count, exceed these values, it may be necessary to either allocate more memory to the job by increasing Max Memory, or reduce the Row Limit to a smaller value. Conversely, if the masked rows are quite small and have few masking assignments, increasing the Row Limit may improve job performance. Remember to consider the worst case (the largest rows, the most masking assignments) table or file format in the Rule Set when making this determination.

- i. **Min Memory (MB)** — Minimum amount of memory to allocate for the job, in megabytes.
- j. **Max Memory (MB)**— Maximum amount of memory to allocate for the job, in megabytes.

Info

It is recommended that the **Min/Max Memory** should be set to at least to **1024**.

- k. **Update Threads**— The number of update threads to run in parallel to update the target database.

Warning

Multiple threads should not be used if the masking job contains any table without an index. Multi-threaded masking jobs can lead to deadlocks on the database engine. Multiple threads can cause database engine deadlocks for databases using T-SQL. If masking jobs fail and a deadlock error exists on the database engine, then reduce the number of threads.

- l. **Nonconforming Data behavior**

- i. **Stop job on first occurrence**- (optional) To abort a job on the first occurrence of non-conformant data. The default is for this checkbox to be clear.

Info

The job behavior depends on the settings specified in the **Algorithm Settings** page and on the individual algorithm pages that define how you view the presence of Nonconforming data. The setting on the **Algorithm Settings** page is global that can be overridden by the setting on the algorithm page for that algorithm. These settings declare if the presence of

Nonconforming data is a failure, or a success for the job. If **Mark job as Failed** is selected as a result of the above settings then the job would be aborted on the first occurrence of nonconforming data. If **Mark job as Succeeded** is selected as a result of the above settings then the job will not be aborted.

- m. **Commit Size** — (optional) The number of rows to process before issuing a commit to the database.
- n. **Feedback Size**— (optional) The number of rows to process before writing a message to the logs. Set this parameter to the appropriate level of detail required for monitoring your job. For example, if you set this number significantly higher than the actual number of rows in a job, the progress for that job will only show 0 or 100%.

Info

Some built-in connectors support the **Disable Constraints**, **Disable Triggers**, and **Drop Indexes** features (see the [Data Source Support](#) page). For built-in connectors implemented using driver support plugins, these options are available via the **Enable Tasks** button. For a full list of built-in connectors using driver support plugins, see [Built-in Driver Supports](#)). For all other built-in connectors, these features will appear as checkboxes.

- o. **Disable Constraints** — (optional) Whether to automatically disable database constraints. The default is for this checkbox to not be selected and therefore not perform automatic disabling of constraints. For more information about database constraints see [Enabling and Disabling Database Constraints](#).
 - p. **Disable Trigger** — (optional) Whether to automatically disable database triggers. The default is for this checkbox to not be selected and therefore not perform automatic disabling of triggers.
 - q. **Drop Indexes** — (optional) Whether to automatically drop indexes on columns which are being masked and automatically re-create the index when the masking job is completed. The default is for this checkbox to not be selected and therefore not perform automatic dropping of indexes.
 - r. **Enable Tasks** - (optional) When this button is pressed, it displays a form with checkboxes next to each task implemented by the driver support plugin being used. The default is for each checkbox to not be selected and therefore not perform any of the tasks. If the masking job being created is for a built-in connector with a builtin driver support plugin, the options displayed will be Disable Constraints, Disable Triggers and Drop Indexes. For a full list of supported built-in connectors and information on specific built-in driver support plugins, see [Built-in Driver Supports](#).
 - s. **Batch Update** — (optional) Enable or disable whether the database load phase to output the masked data will be performed in batches or not. The size of the batches is determined by the **Commit Size** field value. This option is recommended because it typically improves the performance of the masking job.
 - t. **Prescript** — (optional) Specify the full pathname of a file that contains SQL statements to be run before the job starts, or click **Browse** to specify a file. If you are editing the job and a prescript file is already specified, you can click the **Delete** button to remove the file. (The Delete button only appears if a prescript file was already specified.) For information about creating your own prescript files.
 - u. **Postscript** — (optional) Specify the full pathname of a file that contains SQL statements to be run after the job finishes, or click **Browse** to specify a file. If you are editing the job and a postscript file is already specified, you can click the **Delete** button to remove the file. (The Delete button only appears if a postscript file was already specified.) For information about creating your own postscript files see [Creating SQL Statements to Run Before and After Jobs](#)
 - v. **Comments** — (optional) Add comments related to this masking job.
 - w. **Email** — (optional) Add e-mail address(es) to which to send status messages.
3. When you are finished, click **Save**.

Environment							
Name	test1						
Purpose	Mask						
Application Name	test1						
Approval workflow	Disabled						
Job ID ▾	Name	Rule Set	Completed	Status	Action	Edit	Delete
1	 ProfileJob	fileconnector	2021-10-04 09:32	★ Succeeded			
Environments Monitor Settings Admin Audit							D E L P H I X

Enabling and disabling database constraints

Depending on the type of target database you are using, the Delphix Engine can automatically enable and disable database constraints.

The ability to enable and disable constraints ensures that the Delphix Engine can update columns that have primary key or foreign key relationships. You can set Delphix to handle constraints automatically by enabling the **Disable Constraints** checkbox on a Masking job. If the built-in or extended connector is using a driver support plugin, **Disable Constraints** can be enabled via **Enable Tasks**. For a full list of supported built-in connectors and information on specific builtin driver support plugins, see [Built-in Driver Supports](#).

 Delphix does not support the enable/disable constraints feature for all databases. To see which databases are supported, see the [Data Source Support](#) page.

Creating SQL statements to run before and after Jobs

When you create a masking job or a certification job, you can specify standard, static SQL statements to run before (prescript) you run a job and/or after (postscript) the job has completed. For example, if you want to mask a column that has a foreign key constraint to another table, you could use a prescript to disable the constraint and a postscript to re-enable the constraint.

You create prescripts and postscripts by creating a text document with the SQL statement(s) to execute. If the text file contains more than one SQL statement, each statement must be separated by a semicolon [;]. For example to remove records with date_column before December 12th, 2017 before masking a table (owner.table), one would create a prescript file containing the following and associate the prescript file to the masking job that includes the table in its ruleset:

```
DELETE FROM owner.table WHERE date_column < '20171207';
```

Database-specific, SQL programming extensions (such as PL/SQL and Transact-SQL) and dynamic SQL statements are not supported in prescripts and postscripts. However, you can create procedures and functions using your database tooling of choice and call them using standard SQL statements from a prescript or postscript.

Managing jobs

Managing jobs from the environment overview screen

Submitting a job

To submit or resubmit a job from the Environment Overview screen, click the Play icon in the Action column for the desired job.

Upon submitting the job, the masking engine will check if there are enough resources allocated to simultaneously running jobs to determine whether to run or queue the submitted job. There are two resources that the submitted job will be verified against.

1. Maximum memory for all running jobs.
 - This limit defaults to a dynamic calculation of 75% of the entire system's available memory minus 6GB, which is reserved for the masking web application. This calculation can be manually overridden by setting the general application setting `MaximumMemoryForJobs`. To revert a manually overridden limit back to the dynamically calculated limit, set the `MaximumMemoryForJobs` to 0.
2. Maximum number of simultaneously running jobs.
 - This limit defaults to 7 simultaneously running jobs. However, this default value can be overridden by setting the general application setting `NumSimulJobsAllowed` to a different value. The engine also provides a dynamic limit for this resource, which takes the number of available cores on the system minus 1, reserved for the masking engine. This dynamic limit can be used by setting `NumSimulJobsAllowed` to 0.

 If the submitted job causes all of the currently running jobs to exceed either of those limits, the job will be queued and run at a later time when enough of the other jobs stop running to free up resources. To view the the position of the job in the queue, navigate to the [Monitor Screen](#).

Stopping a Job

The Play icon changes to a Stop icon while the job is RUNNING OR QUEUED.

To stop a RUNNING or QUEUED job from the Environment Overview screen:

1. Locate the job you want to stop.
2. In the job's **Action** column, click the **Stop** icon.
3. A popup appears asking, "Are you sure you want to stop job?" Click **OK**.
4. When the job has been stopped, its status changes to CANCELLED.

Stopping a RUNNING job can result in corrupted or semi-masked data. Stopping a QUEUED job will have no impact on the data source, since the execution of the job has not yet begun. If email notifications are enabled, stopping a QUEUED job will send an email to the user who created the job indicating that it has been cancelled by the user who stopped the job.

Verifying a Job

When the job is complete, the status will change to either SUCCEEDED or FAILED.

After the job completes successfully, return to the Inventory and check that the Domain and Method populated automatically for sensitive data. Sample screenshot below.

Home > Environments > env_Y8FEODQ1 > Inventory > con_ETOI9UMT

con_ETOI9UMT

Filter By: All Fields Masked Fields Auto User

Select Rule Set: con_ETOI9UMT

Filter Contents: Search By Name Search Alphabetically

Formats: LZM73M2C.fmt

Type	ID	Position	Method	Domain	Algorithm	Edit	Delete
ASCII	DATA_00	11	Mask	NULL_SL	plg_HOUV1KC...e Increment		
ASCII	DATA_01	61					
ASCII	ID	1					

Monitoring masking job

This section describes how users can monitor the progress of a masking job.

Monitoring masking job refers to the job status or completion state. To determine if the masking operation is completed, you must compare the number of rows in the table to the number of rows masked. If the two are equal, then the masking operation is completed. But this does not indicate that the data masking operation was successful. If the masking script is incorrect, the masking operation may still complete but not produce the desired masked data outcome. To determine whether the data is properly masked, you must perform an audit of a statistical data sample.

Monitoring your masking jobs

Once a masking job has been created and started, you can monitor its progress by navigating to the Monitor tab or by clicking on the name of the masking job on any screen. The monitoring tab shows you a list of executed masking jobs, their progress as well as their current status. To get even more detail on the progress of an individual masking job, click on the Job Name.

The screenshot displays the 'Monitor' page in the DELPHIX Continuous Compliance application. At the top, there's a navigation bar with 'Monitor' selected. Below it, a breadcrumb trail shows 'Home > Monitor'. The main area features a summary of '2/7 Jobs Running' and '2048/6041 Memory Usage (MB)'. A search bar allows filtering by 'Start Date', 'End Date', 'Any Status', and 'All Jobs'. Below the search bar is a table of masking jobs:

Environment	Job Name	Submit Time	Start Time	Type	Progress	Status	Queue Position
TestEnv	m_psg_2	2022-04-25 11:54	2022-04-25 11:54	MASK	3 of 10 - Preparing ...	RUNNING	
TestEnv	m_psg_1	2022-04-25 11:54	2022-04-25 11:54	MASK	3 of 10 - Preparing ...	RUNNING	
TestEnv	p_psg_2	2022-04-25 11:54	2022-04-25 11:54	PROFILE	5 of 5 - Profiling F..	SUCCEEDED	
TestEnv	p_psg_1	2022-04-25 11:54	2022-04-25 11:54	PROFILE	5 of 5 - Profiling F..	SUCCEEDED	

At the bottom of the table, there's a pagination control showing 'Showing 1-4 of 4' and a breadcrumb trail: 'Environments | Monitor | Settings | Admin | Audit'. The DELPHIX logo is visible in the bottom right corner.

Search

On the **Monitor** screen, you can perform a search to view the status of all the masking jobs executed. Use start date, end date, any status, all types, and search bar field to filter the search query as per your requirement.

The filter parameters are:

- **Execution Start Date:** You can select a range for start date using the start date and End date field.
- **Execution Status:** You can filter the result based on execution results like SUCCESS, FAIL, CANCELED and so on.
- **Job Type:** You can filter the execution based on Job Type like Mask, Profile, Tokenization, and Restore.
- **Job Name:** You can apply a wild card search by adding * after the job name using the text field.

Event Status

The following table lists the states of a masking job/event.

Job Status Icon	Status	Description
<input type="checkbox"/>	Cancelled	It appears when a user cancel a running task/execution.
<input type="checkbox"/>	Failed	It appears when there is an error in the execution of an event in the task/execution.
<input type="checkbox"/>	Queued	It means the events of the task/execution are yet to start.
<input type="checkbox"/>	Running	It appears when the event is in progress.
<input type="checkbox"/>	Succeeded	It means the event is completed successfully.It appears when the event is successful
<input type="checkbox"/>	Skipped	It appears when the event is skipped and process moved to the next event of the task/execution.
<input type="checkbox"/>	Non-Conformant	It appears when an event of the task/execution is successful but with warning.

i If a job gets canceled or failed in between then the rest of the running or queued events will be marked as failed or canceled. Except for last event, i.e **Execution Finished** or **Profiling Finished**.

Events

The monitoring tab shows you a list of executed masking jobs along with their current status. To get even more detail on the progress of an individual masking job, click on the Job Name to view more detailed information about the job/event. This screen also displays the sequence in which the events are executed. Click **Execution Logs** if you want to view the log status of an event.

The events are executed in the following sequence:

1. **Init Execution:** The execution has begun.
2. **Collecting Job Configurations:** Collect the Job details stored in the MDS/DB.
3. **Preparing Execution:** Create transformation XML for a kettle that includes, pre-script, post-script, create and drop identity XML.
4. **Execute Pre Execution Custom Driver Task:** Execute custom drive pre-execution tasks.
5. **Start Execution:** Starts the masking Job
6. **Pre Sql Script:** execution prescript if available
7. **Post-Sql Script:** execution postscript if available
8. **Execute Post Execution Custom Driver Task:** Execute post-SQL operations from the custom driver
9. **Collect Job Information:** Collect and store execution information in the database.
10. **Execution Finished:** The execution is finished and removed from monitoring

Queue Position

Queue position refers to the job's numerical order of when it will be dequeued and run, relative to other queued jobs. If a job is not in the queue, it will not have a queue position.

Monitoring a single job

In addition to viewing high-level stats about the status/progress of all your jobs, you can also deep dive into each job to get more details. By clicking the name of the masking job, you will be redirected to a screen with more granular information including; environment name, connector name, job start time, previous run time, number of tables defined in the job, number of jobs tables masked, number of tables to be masked, the type of job, the total time the job has taken, rows remaining to mask, rows masked, number of streams, etc.

Home > Monitor > Completed

Monitor


 DATABASE


 SUCCESS

0
 Jobs Running

p_psg_2

<p>Job Type Profile</p> <p>Environment TestEnv</p> <p>Job ID 4</p> <p>Execution ID 16</p> <p>CM Connection table</p> <p>Source / Target - / hercules</p> <p>Profiling Report</p> <p>4_16.log Log</p> <p>Execution Logs</p>	<ul style="list-style-type: none"> ✔ Init Profiling ✔ Collecting Job Configurations ✔ Preparing Profiling ✔ Start Profiling ✔ Profiling Finished 	<p>Start Time 11:54:01</p> <p>Previous Run Time 00:00:01</p> <p>Total # of Tables 15</p> <p>Tables Profiled 15</p> <p>Tables to be Profiled 0</p> <p>Total Time Taken (HH:mm:ss) 00:00:03</p> <p>Streams 1</p>
--	---	---

Completed Processing Waiting Results

Completed

15
 Complete

15
 Total Tables

ID	Name	Progress	Status/Logs	Total Time (HH:mm:ss)	Rows Per Min	Rows Profiled
192	test1	100%	✔	00:00:01	0	0
195	test10	100%	✔	00:00:01	0	0
182	test11	100%	✔	00:00:02	0	0
185	test12	100%	✔	00:00:02	0	0
189	test13	100%	✔	00:00:02	0	0
186	test14	100%	✔	00:00:02	0	0
184	test15	100%	✔	00:00:02	0	0
190	test2	100%	✔	00:00:02	0	0
187	test3	100%	✔	00:00:02	0	0
181	test4	100%	✔	00:00:02	0	0
194	test5	100%	✔	00:00:02	0	0
188	test6	100%	✔	00:00:02	0	0
183	test7	100%	✔	00:00:02	0	0
193	test8	100%	✔	00:00:02	0	0
191	test9	100%	✔	00:00:02	0	0

In addition to seeing this additional information about each masking job, you can look into the status/progress of each table/file defined in the masking job. Each table/file will be separated into 1 of 4 tabs:

- **Completed:** The Completed tab shows which tables or files the job has completed and includes information such as the rows masked per minute, rows masked, and rows remaining.
- **Processing:** The Processing tab will include information on the tables or files the job is currently processing.
- **Waiting:** The Waiting tab shows us which table or files are waiting to be processed.
- **Results:** The Results tab shows a list of tables that are masked.

Displaying non-conformant data

When non-conformant data is encountered by a masking job, the job will either Fail or Succeed with a warning, depending on how the algorithms associated with the ruleset for the job are configured. As depicted in the screenshot, the non-conformant data can be accessed via the **Completed** tab on the Monitor page for the job, which can be accessed by clicking on the Job name from the Environment Overview page. In the main body of the Monitor page, a summary of the **Tables with Nonconforming Data** and **Columns with Nonconforming Data** is reported. Further details on the non-conformant data encountered can be accessed by clicking the Success or Fail icon next to each table or file listed in the **Completed** tab.

Success Report - testdata_XML ✕

NONCONFORMING DATA Learn More			
Event	Cause	Approximate Row Count	Description
UNMASKED_DATA	PATTERN_MATCH_FAILURE	1000	Column RCHARS64_T1_0 contained nonconforming data that was not masked by algorithm DateShiftVariable The top nonconforming data samples were: LLLLLL LLLLLL LLLL LLLLLL LLLLLLLLLL LLLLLLLLLL LLLLLL
1_27_testdata_XML.txt			
<pre> 2019/03/12 21:04:04 - testdata_XML - Loading transformation from XML file [/var/delphix/masking/output/Test/DMSApplicator/Oracle/1/KETTLE_MASK_XML_1_testdata_XML_27.xml] 2019/03/12 21:04:04 - testdata_XML - Using legacy execution engine 2019/03/12 21:04:04 - KETTLE_MASK_XML_1_testdata_XML_27 - Dispatching started for transformation [KETTLE_MASK_XML_1_testdata_XML_27] 2019/03/12 21:04:05 - Table input.0 - Finished reading query, closing connection. 2019/03/12 21:04:05 - Select values.0 - Finished processing (I=0, O=0, R=1000, W=1000, U=0, E=0) 2019/03/12 21:04:05 - Get All Lookups Values.0 - Finished processing (I=0, O=0, R=1000, W=1000, U=0, E=0) 2019/03/12 21:04:05 - Table input.0 - Finished processing (I=1000, O=0, R=0, W=1000, U=0, E=0) 2019/03/12 21:04:06 - User Defined Java Class.0 - Finished processing (I=0, O=0, R=1000, W=1000, U=0, E=0) 2019/03/12 21:04:06 - SelectValues_MetaData.0 - Finished processing (I=0, O=0, R=1000, W=1000, U=0, E=0) 2019/03/12 21:04:06 - String Cut.0 - Finished processing (I=0, O=0, R=1000, W=1000, U=0, E=0) 2019/03/12 21:04:07 - Update.0 - Finished processing (I=1000, O=0, R=1000, W=1000, U=1000, E=0) </pre>			

The non-conformant data events are displayed followed by the masking log for the table or file. If there were no non-conformant data events, "None" is displayed under **NONCONFORMING DATA**, otherwise, for each type of non-conformant data, a row will be displayed with the following information:

- **Event type:** either JOB_ABORTED or UNMASKED_DATA if the job was not aborted.
- **Cause:** always PATTERN_MATCH_FAILURE.
- **Approximate Row Count:** approximate number of rows with non-conformant data (at least within an order of magnitude).
- **Description:** details the name of the column or field with non-

Interpreting samples of non-conformant data patterns

Each character in the non-conformant data is sampled per its [Unicode Character Property](#).

- N for digits
- L for letters
- M for marks
- P for punctuation
- S for symbols
- Z for separator
- O for other
- U for unknown

Tracking Non-conformant Data

 Please note that actual personal data is never displayed, only the samples (a.k.a. patterns) of non-conformant data are displayed on this page

Using the DataBase specific SQL query, it is possible to locate data corresponding to the non-conformant data sample. The table and column names can be found on the table report. In the example above, the table name is "testdata_XML" and the column name is "RCHARS64_T1_0".

 **Note**
The pattern might be not an exact representation of the data in the field, but a part of the data. For instance, white spaces at the beginning or at the end of the data might be truncated.

Oracle DB specific example

Below are the [Oracle character classes](#), used in the regular expression:

Character Class Syntax	Meaning
[[:alnum:]]	All alphanumeric characters
[[:alpha:]]	All alphabetic characters
[[:blank:]]	All blank space characters.
[[:cntrl:]]	All control characters (nonprinting)

Character Class Syntax	Meaning
[digit:]	All numeric digits
[graph:]	All [punct:], [upper:], [lower:], and [digit:] characters.
[lower:]	All lowercase alphabetic characters
[print:]	All printable characters
[punct:]	All punctuation characters
[space:]	All space characters (nonprinting)
[upper:]	All uppercase alphabetic characters
[xdigit:]	All valid hexadecimal characters

For the LLLLL sample in the example above, Oracle DB SQL query would look like:

```
SELECT RCHARS64_T1_0 FROM testdata_XML WHERE regexp_like(RCHARS64_T1_0, '[:alpha:]{5}');
```

For the LLLLZLLLZLLL sample, the Oracle DB SQL query would look like:

```
SELECT RCHARS64_T1_0 FROM testdata_XML WHERE regexp_like(RCHARS64_T1_0, '[:alpha:]{4}[:space:][:alpha:]{3}[:space:][:alpha:]{4}');
```

Limitation for the multi-column extensible algorithm

If a Non-conformant data pattern is encountered - it is displayed for all the masked columns of the MC Algorithm, not only for the column where that event has occurred. In that case, the manual analysis of the error message will be required to find the actual column(s) with the Non-conformant data.

Masking job wizard

The Continuous Compliance job wizard enables users to create and modify masking jobs. While the wizard facilitates a number of workflows and operations, more advanced functionality and finer control of features are available directly in the masking application. The Job Wizard currently functions only with certain data platforms, but these constraints do not apply when working directly in the masking application.

Supported data platforms

The following data platforms are currently supported from within the Job Wizard: - Oracle Database - RDS Oracle Database - MSSQL Server Database - Sybase Database

This restricted list only affects your use of the wizard; an expanded number of platforms are supported directly in the masking application. Some operations within the Job Wizard are also limited. See below for details.

Supported operations

While creating a masking job in the Job Wizard, you are able to do the following:

- Create a new application or use an existing application
- Create a new environment or use an existing environment
- Create a new connector
- Create a new rule set
- Update inventory
- Create a masking job
- Update a masking job
- Change the connector for an existing job
- Change the rule set for an existing connector
- Run a newly created job immediately
- Run an updated job immediately after the update

 Operations marked with an asterisk are limited in the Job Wizard but fully supported in the main application.

What is not supported in the wizard

The following data platforms and operations are not supported in the Job Wizard. To access additional functionality, use the main masking application.

Unsupported data types

The following data types are supported when using the main masking application but are not currently supported in the Job Wizard:

- DB2 Database
- PostgreSQL Database
- Generic Database
- Delimited File
- Excel Sheet File
- Fixed File
- Mainframe Data Set

- XML File

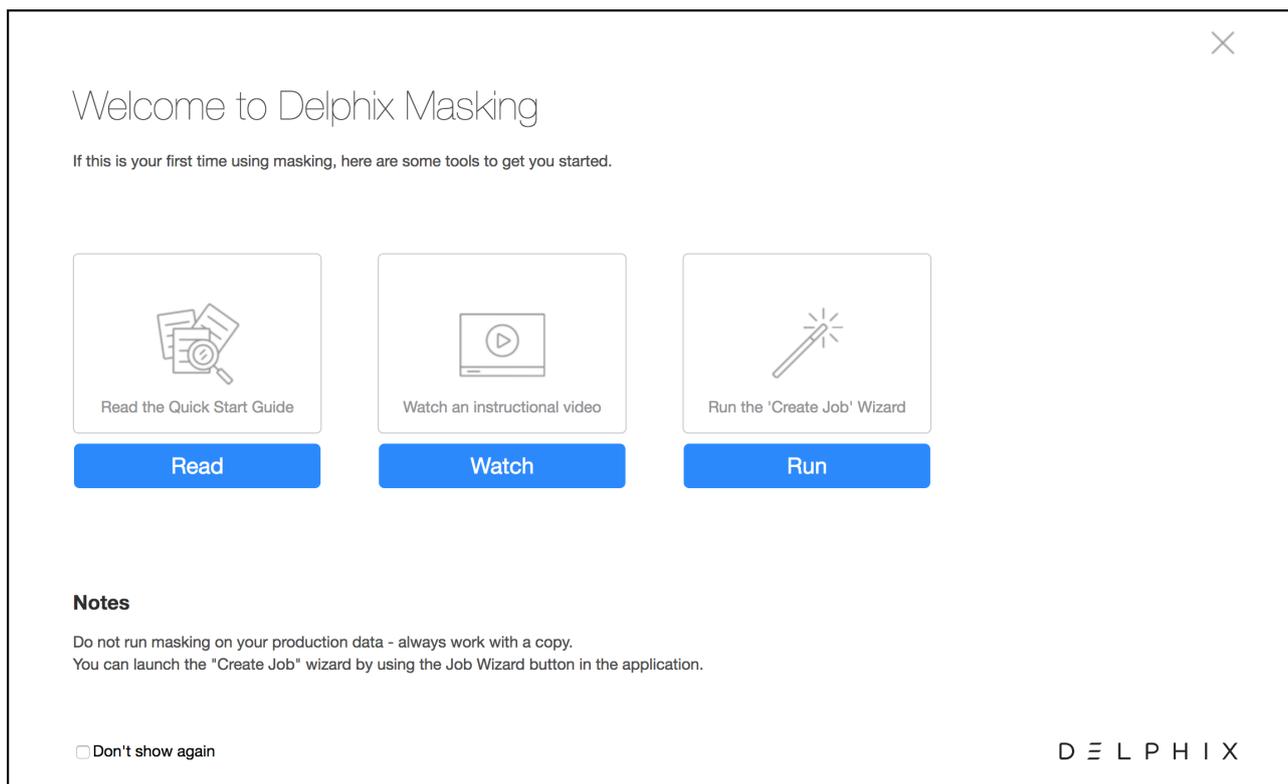
Unsupported operations

The following operations are not yet supported from within the Job Wizard:

- Creating any connector or rule set for an unsupported data type
- Deleting any application, environment, connector, rule set, or masking job
- Importing or exporting any object
- Updating an environment
- Creating a connector using Advanced mode
- Updating a connector
- Updating a rule set
- Creating a job for an unsupported data type
- Modifying a job for an unsupported data type
- Monitoring running jobs
- Creating, editing, deleting, or running any Profile jobs

Opening the masking job wizard

When you first login to masking, the welcome screen offers a link to learn more or begin masking immediately. To open the Job Wizard, click Run on the welcome page.



To use the Job Wizard from the masking application, click the Create Job button in the upper right-hand corner, as highlighted in the screenshot below.

DELPHIX MASKING Job Wizard admin

Environments Monitor Settings Admin Audit

Home > Environments Select Action

Environments

Search Search

Environment ID	Application ▲	Environment	Purpose	No of Jobs	Edit	Export	Copy	Delete
4	test	Test_MASK	Mask	0				
3	test	Tokenization Re...	Tokenize/Re-Identify	0				
1	test1	test1	Mask	1				

[Go to top of page](#)

[Environments](#) | [Monitor](#) | [Settings](#) | [Admin](#) | [Audit](#) D E L P H I X

⊞ Only administrators or users with the following privileges can see the Create Job button.

- Environment: View, Add, and Update
- Connection: View and Add
- Rule Set: View and Add
- Inventory: View and Update
- Profile Job: View, Add, and Run
- Masking Job: View, Add, Update, and Run
- Inventory Report: View

Creating a new masking job

The Job Wizard makes creating a new masking job much easier by guiding you through the process. You can create new objects or choose to use existing ones that have already been defined. When creating a new masking job, the Job Wizard follows this sequence:

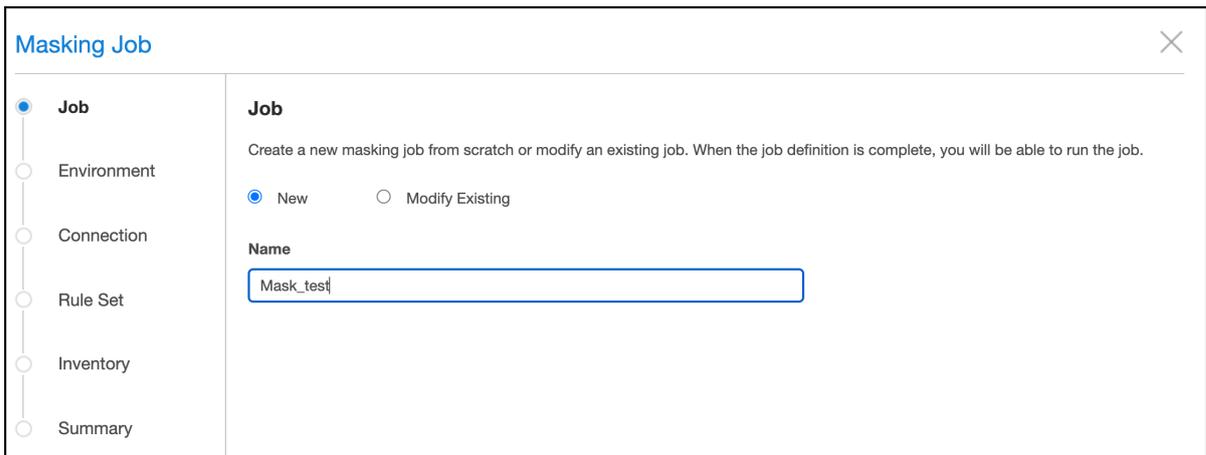
- Job Naming
- Application/Environment Selection
- Connection Selection
- Rule Set Selection
- Inventory Selection
- Summary Page

You can navigate back and forth through the pages of the Job Wizard.

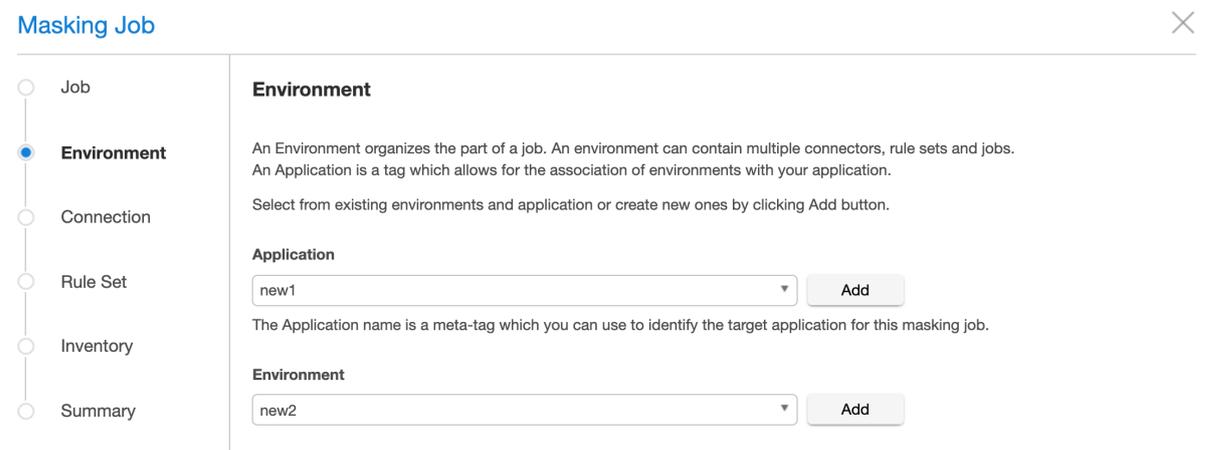
⊞ If the product times out due to long inactivity, you will need to start over.

To create a new masking Job using the new Job Wizard, follow the procedure below:

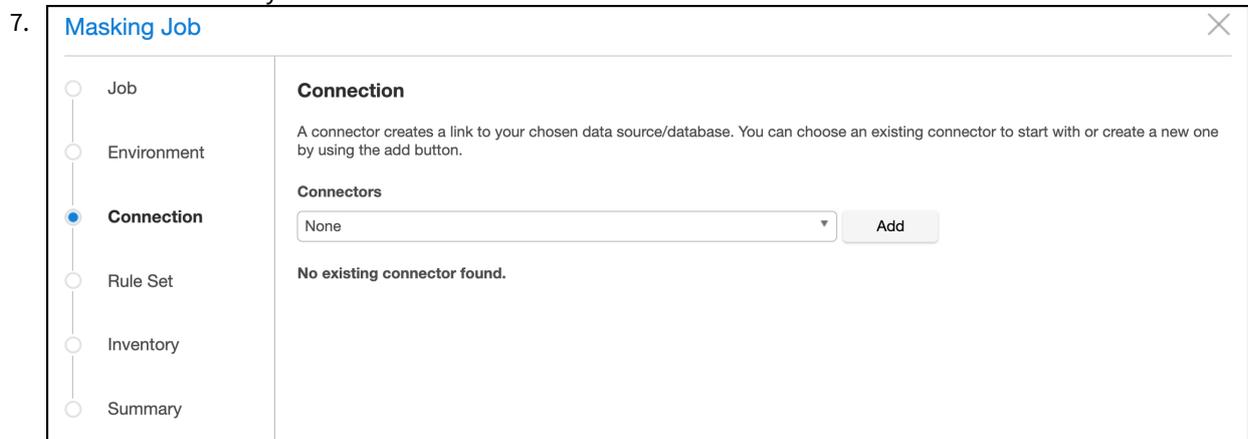
1. Log into your Continuous Compliance Engine and from the Welcome screen select Run.
2. Select the New radio button and enter a name for your Masking job.



3. Click Next.
4. From the drop-down menu select an Application and Environment. If none exist use the Add button to add one.



5. Click Next.
6. Select a Connector from the drop-down menu. If none exists select the Add button, then use the Add Connector dialog to add a new connector. The Job Wizard only supports the following Connector types:
 - Database - MS SQL
 - Database - Oracle
 - Database - RDS Oracle
 - Database - Sybase



Click Next.

- On the Rule Set screen select an existing Rule Set or create a new one by clicking the Add button.

Masking Job

Rule Set

The rule set is the set of data to which you want to secure with masking. You can begin by using an existing rule set or you can create a new one by clicking on the Add button.

Changes that you make on the screen will be saved when you click Next.

Note that the size of the database and connection speed can impact load time.

Rule sets:

test

Total selected Tables: 0 **View:** All

Table Name
<input checked="" type="checkbox"/> DBVERIFICATION_TABLE
<input type="checkbox"/> Foo
<input type="checkbox"/> foo_test
<input type="checkbox"/> Foo1
<input type="checkbox"/> foo1010101
<input type="checkbox"/> foo111122
<input type="checkbox"/> foo121333

Showing 1-10 of 16 tables

10 tables per page

Cancel Back **Next** Save Job

- Click Next.
- From the Inventory screen select how your data will be masked. In the screenshot below we are masking subscriber last names.

Masking Job



- Job
- Environment
- Connection
- Rule Set
- Inventory**
- Summary

Inventory

The inventory defines how your data will be masked. Choose a domain for the column which you wish to mask, and then select the algorithm you would like used for the mask.

Delphix Masking has an automatic masking Profiler that you can run separately, which will scan data in the Rule Set to identify sensitive data and will automatically apply default algorithm. Refer to [How to setup your Inventory](#) for more information.

Note that the size of the database and connection speed can impact load time

Rule Set : test

View All columns Apply

Tables ▲	Column	Masking Settings
Foo	IDD	Data type int(0) Domain NO MASKING Algorithm NONE Automatic Profile Overwrite <input checked="" type="checkbox"/> Notes <input type="text"/> Clear

Showing 1-1 of 1 Tables Showing 1-1 of 1 Columns

Cancel Back Next Save Job

11. Click Next.
12. The final screen of the Job Wizard displays a Summary of your selections.

Masking Job

Summary

Job

Name	Mask_test
Application	new1

RuleSet

Name	test
------	------

Environment

Name	new2
Purpose	Masking

Inventory

Number of tables	0
Number of columns	0

Connection

Name	test_mysql
Data Source Type	Database
Database name	mssql
Schema	Biscuit
Host name/IP address	mgu-mysql-new.dlpxdc.co
Port	1433
User login	sa

Run Masking Job Now and go to Monitor progress

Cancel Back Next Save Job

- Clicking Run Masking Job Now and go to Monitor progress, saves your job, and runs it immediately. Save Job allows you to save your job and run it at a later date. Note: Selecting this option means your data will not be masked until you run the job.

When objects are saved

Application, environment, connector, and Rule Set objects are created and persist after you click the Add button and see a success message. If you cancel the Job Wizard before completing the job setup, the objects you created will be saved, and they will be available for use the next time you launch the Job Wizard.

The Inventory definition is saved when you change the selection of a table or column, or when another View filter is applied.

The masking job is saved when you click either Save Job or Run Masking Job Now and go to Monitor progress and a success message is returned on the Summary screen.

Updating an existing masking job

You can use the Job Wizard to modify any masking job that targets a supported data type.

- On the Job screen of the Job Wizard, select Modify Existing
- From the list of available jobs select the one you want to modify. This list only shows jobs that are supported by the wizard. You can filter the job list by selecting the filter icon.
- Once you select a job, you can change the following as part of the Modify flow:

- Change/create a new Connector
- Change/create a new Rule Set
- Update inventory
- Save or run the modified job

You cannot alter application and environment settings as part of the Modify flow, but you can do so in the main masking application.

Running stopping jobs

Running and stopping jobs from the environment overview screen

To run or rerun a job from the Environment Overview screen:

- Click the Run icon (play icon) in the Action column for the desired job.

The Run icon changes to a Stop icon while the job is running. When the job is complete, the Status changes.

To stop a running job from the Environment Overview screen:

1. Locate the job you want to stop.
2. In the job's Action column, click the Stop icon.
3. A popup appears asking, "Are you sure you want to stop job?" Click OK.
4. When the job has been stopped, its status changes.
5. After the job completes successfully, return to the Inventory screen and check that the Domain and Method populated automatically for sensitive data. Sample screenshot below.

Type	ID	Position	Method	Domain	Algorithm	Edit	Delete
ASCII	DATA_00	11	Mask	NULL_SL	plg_HOUV1KC...e Increment		
ASCII	DATA_01	61					
ASCII	ID	1					

Masked provisioning

This section contains the following topics:

- [Configuring virtualization service for masked provisioning](#)
- [Provision masked VDBs](#)

Configuring virtualization service for masked provisioning

Introduction

During the VDB provisioning process, the Virtualization Engine can optionally run a masking job from a Continuous Compliance engine on the VDB. Use these instructions to customize the host address, port number, and/or login credentials that the Virtualization Engine will use to contact the Masking Engine.

Important validation notices

When configuring masked provisioning, ensure that the versions of the Virtualization Engine and Masking Engine are compatible. See the [compatibility matrix](#).

Old versions of the serviceconfig or any information associated with them are not tracked. In particular, if you have been using the local masking service or a remote service and then change to a new remote service Delphix will start throwing out any old job information on the next masking job/fetch or GUI reload. Users should not rely on that information being preserved through serviceconfig updates.

Delphix does not validate network availability between the two engines or any other hosts that both engines might want to communicate with. The state or availability of either host is not checked, if either host becomes unduly slow, congested, or unresponsive Delphix will not be able to issue compelling warnings regarding those issues.

Instructions

Use these instructions to customize the host address, port number, and/or login credentials that the Virtualization Engine will use to contact the Masking Engine.

-  This does not alter the Continuous Compliance Engine UI port. It is specific to coordinating communication between the Virtualization Engine and a Masking Engine about available masking jobs and job results.

To change the Virtualization Engine's connection details for its Masking Engine:

1. Using a shell, login to the **CLI** using:
 - On 5.2 and earlier releases: **delphix_admin**.
 - On 5.3 and later releases: **admin**.
2. At the **CLI** root prompt, type **maskingjob**.
3. At the **maskingjob** prompt, type **serviceconfig**.
4. To list service configurations, type **ls**.
5. At the serviceconfig, type **select `MASKING_SERVICE_CONFIG-1**.
6. To view the configurations, type **ls**.
7. With this service config selected, enter **update**.
8. In the update mode, use the **set** command to modify the configuration. For example, type `set port=[YOUR DESIRED PORT NUMBER]` to change the port number.
9. Commit the change by typing **commit**.
10. Type **ls** to confirm the configurations.
11. Type **exit** to exit the CLI.

Provision masked VDBs

Masked virtual databases (VDBs) function just like normal VDBs. The only distinction is that the data they contain has been masked by a masking job. Masked VDBs can be replicated to a separate Delphix Engine (in non-prod) without sending the original data that was obfuscated during masking using a process called Selective Data Distribution (SDD). This topic describes how to work with masked VDBs.

Prerequisites

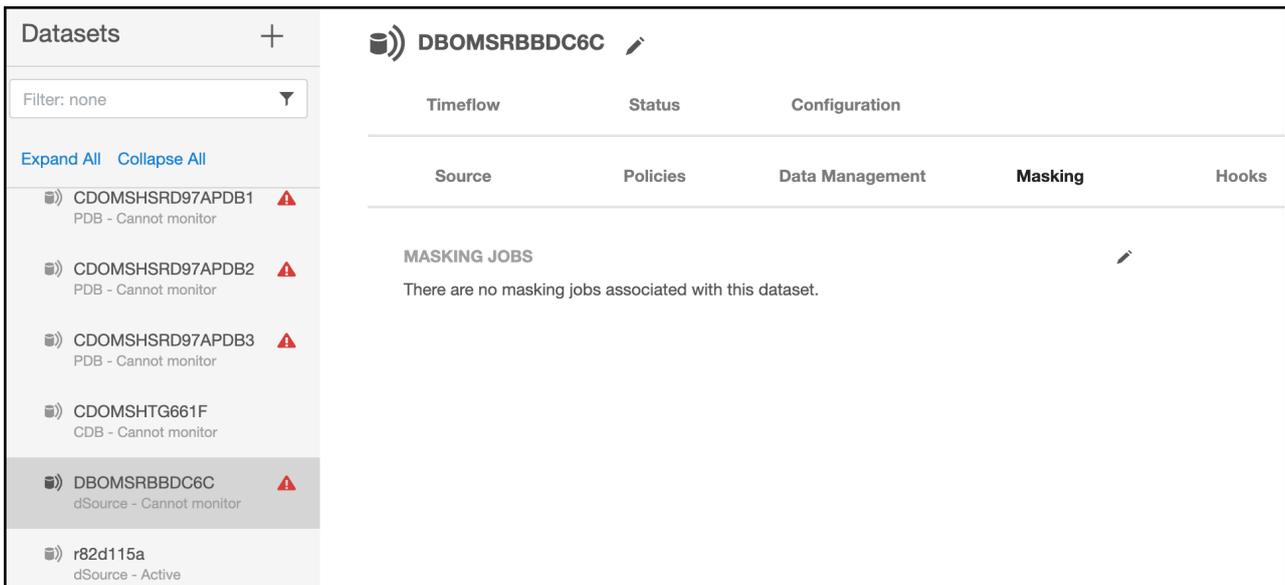
Before attempting to create a Masked VDB, you should be familiar with both Delphix Virtualization and Delphix Masking concepts and workflows.

Restrictions

- A single masking job cannot be assigned to multiple VDBs simultaneously. If you are using the same masking ruleset on multiple VDBs, be sure to create a unique job for each VDB to avoid any issues with provisioning or refreshing.
- Provisioning or refreshing masked VDBs is only supported for Oracle, MS SQL Server, and Sybase. Provisioning or refreshing other types of masked VDBs such as DB2 are not supported.
- You cannot apply additional masking jobs to a masked VDB or its children.
- If a masking job has been applied to a VDB, you cannot create an unmasked snapshot of that VDB.
- Masking must take place during the process of provisioning a VDB. If an existing VDB has not had a masking job applied to it, then you cannot mask that particular VDB at any point in the future. All the data within the VDB and its parents will be accessible if it is replicated using SDD.
- When selecting a connector to use for Masked Provisioning, a "basic" connector must be used **unless** you are masking an Oracle Pluggable Database (PDB), in which case an "advanced" connector must be used.
- Only in-place masking jobs can be selected.
- Masked Provisioning is supported on Oracle RAC only when used with "script-based masking" and not when a masking job is used for SDD.

Identifying and navigating to masked VDBs

Masked VDBs appear in the Virtualization Engine's Datasets pane, just like regular VDBs. They are most obviously identified by the different icons used to represent them. In addition, a masked VDBs Configuration tab will contain information about the masking job that you applied to it. Generally, anything you can do with an unmasked VDB is also possible with a masked VDB.



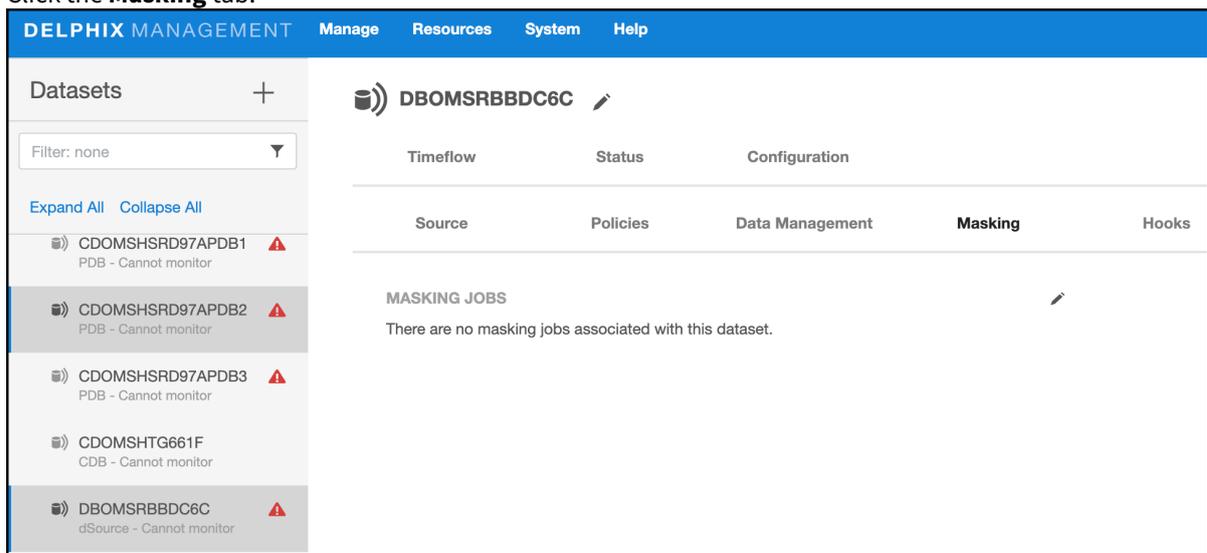
Provisioning masked VDBs

- In the Virtualization Engine, associate a masking job with a dSource.
- Use the dSource provision wizard to provision a VDB with a masking job.

Associating a masking job with the dSource

To provision a masked VDB, you must first indicate that the masking job you are using is complete and applicable to a particular database. You do this by associating the masking job with a dSource.

1. In the **Datasets** panel on the left-hand side of the screen, click the dSource to which the masking job is applicable and with which it will be associated.
2. Click the **Configuration** tab.
3. Click the **Masking** tab.



4. Click the **pencil** icon to edit. All masking jobs on this Delphix Engine that have not been associated with another dSource will be listed on the right-hand side.

5. Select the **job** you want to associate with this dSource.
6. Click the tickmark symbol to confirm.

The screenshot shows the Delphix Engine interface. On the left, the 'Datasets' panel lists several datasets, including 'DBOMSRBDC6C' which is highlighted. On the right, the configuration page for 'DBOMSRBDC6C' is shown, with tabs for 'Timeflow', 'Status', 'Configuration', 'Source', 'Policies', 'Data Management', 'Masking', and 'Hooks'. A 'MASKING JOBS' dialog box is open, displaying a search field for 'Masking Job Name' and a 'No Rows To Show' message. A blue checkmark button is visible in the bottom right corner of the dialog.

7. Repeat for any other jobs that you want to associate with this dSource at this time.

The Delphix Engine now considers this masking job to be applicable to this dSource and ready for use. When provisioning from snapshots of this dSource, this masking job will now be available.

[-] Masking jobs can also be associated with virtual sources in addition to dSources.

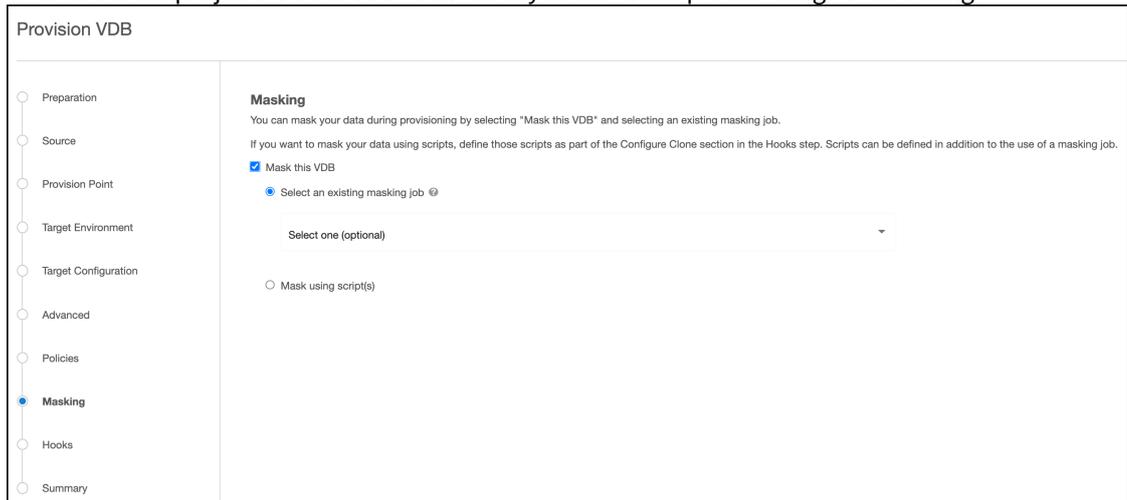
A masking job must be Multi-Tenant for creating a masked VDB. The Multi-Tenant option allows existing rulesets to be reused to mask identical schemas via different connectors. The connector can be selected at job execution time.

Provisioning a masked VDB using the dSource provisioning wizard

The steps required to provision a masked VDB are almost identical to the steps required to provision an unmasked VDB. Once you have created a masked VDB, you cannot unmask it, nor can you alter which masking job it uses. All snapshots in the VDBs TimeFlow will always be masked using the masking method that you selected when you provisioned the masked VDB.

1. In the **Datasets** panel on the left-hand side of the screen, select the dSource.
2. Click the **TimeFlow** tab.
3. Click the **Provision VDB** icon.
4. Review the information for Installation Home, Database Unique Name, SID, and Database Name. Edit as necessary.
5. Review the Mount Base and Environment User. Edit as necessary.
 - If you want to use login credentials on the target environment that are different from the login credentials associated with the Environment User, select Specify Privileged Credentials.
6. Click **Next**.
7. If necessary, edit the **Target Group** for the VDB.
8. Select the **None** option for the Snapshot Policy for the VDB.

- **Snapshot Policy Selection:** For almost all use cases involving Masked VDBs, a Snapshot Policy of None is appropriate. Using a Snapshot Policy in conjunction with SDD can result in the leak of sensitive data.
9. Click **Next**.
 10. Click **Mask this VDB**. You will be presented with two options to mask this VDB:
 - Select an existing masking job: Choose this option if you want to mask using a preconfigured Masking Job. Only masking jobs that have been associated with the parent dSource will be available.
 - **Selecting Unique Masking Jobs:** If you are using the same masking ruleset on multiple VDBs, be sure to create a unique job for each VDB to avoid any issues when provisioning or refreshing.



- Masking using scripts(s): Alternatively, you may define some Configure Clone scripts in the Hooks step to perform masking.
- **Defining Configure Clone Hooks to Mask VDB:** If you choose to mask using script(s), you must define the Configure Clone Hooks to run masking jobs yourself. If you don't define any Configure Clone hooks in the Hooks step, the data will be marked as masked, but it will not be masked.



11. Click **Next**.
12. Specify any **Pre or Post Scripts** that should be used during the provisioning process. If the VDB was configured before running the masking job using scripts that impact either user access or the database schema, those same scripts should also be used here. Be sure to define the **Configure Clone** hooks to run the masking job if you choose to mask using a script(s) in the Masking step.

13. Click **Next**.
14. Click **Submit**.

If you click Actions in the upper right-hand corner, the Actions sidebar will appear and list an action indicating that masking is running. You can verify this and monitor progress by going to the Masking Engine page and clicking the Monitor tab.

Environment	Job Name	Submit Time	Start Time	Type	Progress	Status	Queue Position
env_Y8FEODQ1	MAsk_new	2021-10-08 07:27	2021-10-08 07:27	[Icon]	0%	Running	
env_Y8FEODQ1	mask_VX6KMROG	2021-10-08 07:26	2021-10-08 07:26	[Icon]	0%	Running	

Once you have created a masked VDB, you can provision its masked data to create additional VDBs, in the same way, that you can provision normal VDBs. Since the parent masked VDB contains masked data, child VDBs will only have masked data. This is a great way to distribute multiple independent copies of masked data that is both time and space-efficient.

Refresh a masked VDB

You refresh a masked VDB in exactly the same way as you refresh a normal VDB. As with provisioning a masked VDB, the masking job will be run during the refresh process.

1. Login to the Delphix Management application.
2. Click Manage.
3. Select Datasets.
4. Select the VDB you want to refresh.
5. Click the Refresh VDB button (2 circular arrows).
6. Select More Accurate and Next.
7. Select the desired refresh point snapshot or click the eye icon to choose the latest available range, A point in time, or An SCN to refresh from.
8. Click Next.
9. Click Submit to confirm.
10. Click the Actions link to watch the progress of the refresh job.
11. To see when the VDB was last refreshed/provisioned, check the Time Point on the Status page.

Disassociating a masking operation on a dSource

If a masking job is found to be unsuitable or should be retired, you can disassociate it through the same database card that you used to associate it.

1. Deselect the job.
2. Click the green arrow to confirm. Note that this will only prevent the creation of new masked VDBs with this job. It will not alter existing masked VDBs in any way. When disassociating a job, review the existing masked VDBs and consider whether you need to delete or disable any of them.

Masked VDB data operations

The following data operations are available to masked VDBs:

- **Rewind:** Alter the database to contain masked data from a previous point in time.
- **Refresh:** Get new data from the parent dSource and mask it.
- **Disable:** Turn off the database and remove it from the host system.
- **Enable:** Turn on the database and make it available on the host system.

Continuous Data and Continuous Compliance Engine compatibility matrix

Virtualization Engine Version	Masking Engine Version
5.0 releases	5.0 releases (minor versions do not need to match)
5.1 releases	5.1 releases (minor versions do not need to match)
5.2 releases	5.2 releases (minor versions do not need to match)
5.2.5.0 (or later 5.2 minor release)	5.2.5.0 (or later 5.2 minor release)

Virtualization Engine Version	Masking Engine Version
5.3.0.0 and later, including later major releases (e.g. 6.0)	5.3.0.0 and later, including later major releases (e.g. 6.0) and minor versions do not need to match

Authoring extensible plugins

This section covers the following articles:

- [Introduction \(Authoring extensible plugins\)](#)
- [General plugin structure](#)
- [Setting up your development environment](#)
- [Algorithms \(Authoring extensible plugins\)](#)
- [Driver supports](#)
- [Managing plugins using the API client](#)
- [Installing a plugin onto the Delphix masking engine](#)
- [Secure plugin deployment](#)
- [Terminology](#)

Introduction (Authoring extensible plugins)

The SDK was formerly referred to as the *Masking Algorithm SDK*, but it is now referred to as the *Masking Extensible SDK*, as of SDK version 1.5.0, as it now allows for the development of different types of extensible plugins. As of Delphix release 6.0.3.0, the Delphix Masking Engine supports the installation of plugins, written in Java, that provide new masking algorithms; and as of 6.0.9.0, driver support plugins. The former feature is referred to as Extensible Algorithms and the latter is referred to as Extensible Driver Supports. This section of the documentation details all aspects of masking algorithm and driver support plugin usage and development. The *Guided Tour* portion of the workflows section for [Extensible Algorithms](#) and [Extensible Driver Supports](#) walk the user through the basic process of building a simple plugin and installing it onto the Delphix Masking Engine. Other sections explore in-depth topics such as making algorithms configurable, consuming input files, etc.

This documentation assumes the reader has some familiarity with Java development as well as operation of the Delphix Masking Engine via both the UI and Web API Client. The reader should also understand the security requirements associated with any new algorithms being developed.

Before getting started

This documentation assumes you have a functional Java 8 development environment. Instructions for setting up a basic development environment are [here](#).

You should also download the [Extensible SDK binary package](#) from the Delphix [download site](#) and unpack it into a new directory on your development system. This directory - the root of the unpacked archive - will be referred to as ***sdk_root***.

It's helpful to add the binaries directory to your PATH. On a UNIX like system, this command will add the SDK utilities to PATH:

```
$ PATH=$PATH:$(pwd)/sdkTools/bin
```

It is presumed that the SDK bin directory is in the user's PATH throughout this documentation.

SDK features

The Extensible SDK provides a number of useful functions that aid development of new algorithms and driver supports for the Delphix Masking Engine. It is available on the Delphix software [download site](#).

- Creation of empty "skeleton" projects, with build files - the maskScript *init* sub-command
- Creation of empty class files for algorithms and driver supports - the maskScript *generate* sub-command
- Testing of masking algorithms and driver supports without a masking engine
 - The maskApp CLI (*only algorithms*)
 - The maskScript *mask* sub-command (*both algorithms and driver supports*)
- Uploading of plugins to the masking engine - the maskScript *install* sub-command
- Sample algorithms and driver supports that illustrate the usage of key features of the Masking Plugin API

Versions compatibility

The SDK shares some key elements with the Masking Engine, so in order for the SDK to provide behaviors as close as possible to the Masking Engine, use the SDK version which corresponds to the Masking Engine where you are planning to use the created algorithm(s). The SDK and the Masking Engine use a common Masking API which provides the mechanisms to run the extensible algorithms. Masking algorithms built on the SDK using the latest Masking API will not necessarily run on an older Masking Engine version.

Delphix Release	Masking API*	Extensible SDK*
6.0.3	1.0.0	-
6.0.4	1.1.0	1.0.0
6.0.5	1.1.0	1.1.0
6.0.6	1.2.0	1.2.0
6.0.7	1.3.0	1.3.0
6.0.8	1.4.0	1.4.0
6.0.9	1.5.0	1.5.0
6.0.10	1.6.0	1.6.0
6.0.11	1.6.0	1.6.0
6.0.12	1.7.0	1.7.0
6.0.13	1.8.0	1.8.0



Prior to Delphix Release 6.0.9 and SDK release 1.5.0, Masking API was referred to as Algorithm API and Extensible SDK as Algorithm SDK.

Several other sources of information are available to aid in plugin development:

- The README.md file under docs in the Extensible SDK download archive
- The [Masking Plugin API Javadoc](#)
- Invoke **maskScript** (located under *sdkTools/bin* in the SDK download) with the -h option for usage help
- Type help at the **maskApp** (also under *sdkTools/bin* in the SDK download) command prompt

General plugin structure

This section covers the following topics:

- [Introduction \(General plugin structure\)](#)
- [Dependency management](#)
- [Plugin Metadata](#)
- [Versioning](#)

Introduction (General plugin structure)

This section describes the structure of the plugin Java archive (JAR) files used to extend the Continuous Compliance Engine with additional algorithms. This includes the **MaskingAlgorithm** interface that classes providing new algorithm code must implement, and various other metadata present in the plugin JAR required for the plugin to be usable. It also discusses some aspects of build dependencies and common pitfalls involved when adding new 3rd party dependencies.

Plugins for the Masking Engine should be self-contained. This means they should include all Java classes necessary to run, with a few critical exclusions. The Java classes that comprise the Masking Plugin API itself are the exception; these must be excluded from the plugin JAR to ensure that the plugin properly uses the API classes present on the Masking Engine (or SDK during the test process). This is described in more detail in the [dependency management section](#).

Dependency management

A vast assortment of third-party Java libraries are available, expanding the set of ready-to-use functionality well beyond what is already a rich standard library. Plugins for the Delphix Masking engine are able to make use of external libraries, but a number of guidelines should be followed to ensure proper function and compatibility. Note that the plugin classloader uses a plugin-first loading strategy for dependencies.

How to properly use and embed external libraries

When using an external library in a plugin for the Delphix Masking engine, consider these guidelines:

- The plugin JAR should contain all external libraries. This is commonly referred to as a "fat JAR". This prevents the plugin code from inadvertently linking with copies of the same library that might happen to be part of the Masking Engine's codebase, leading to potential version conflicts and unpredictable behavior across upgrades.
- **However**, a small set of packages defining the interface between plugins and the Delphix Masking Engine **must not** be embedded in the JAR. It is critical that, for these packages, the plugin code link against the same classes already loaded by the engine. These packages are:
 - com.delphix.masking:masking-algorithm-api
 - com.fasterxml.jackson.core:jackson-annotations
 - com.google.code.findbugs:jsr305
 - junit:junit

If the externally created plugin uses any of the mentioned libraries, the exact same libraries versions should be used by the plugin author, as the ones used by the SDK. The way to find those versions is:

```
- in the installed SDK find the following gradle file under `samples` directory:
  * gradle.properties
```

It contains the versions of the SDK provided external libraries, for example:

```
googleGuavaVer=28.0-jre
maskingAlgoVer=1.3.0
jacksonVer=2.9.5
junitVer=4.12
```

Looking to those versions author should decide what version of corresponding library to use (if it is required by their design).

- Plugins consuming third-party libraries should be thoroughly tested, as it is not uncommon that library code will attempt to use permissions not granted by the plugin sandbox. If this is the case, there is currently no way to modify the constraints under which the plugin code is executed.
- The plugin author, not *Delphix*, is responsible for ensuring that any license files or other forms of attribution required by any embedded software are handled properly.
- The entity deploying the plugin, not *Delphix*, is responsible for ensuring the organization operating the Masking Engine has obtained the necessary licenses or rights to use any embedded software.

Example build file

The following fragments, derived from the sample algorithm *build.gradle* file, illustrate how to correctly build a plugin using the gradle build system:

```

jar {
    from {
        configurations.runtimeClasspath.collect { it.isDirectory() ? it :
zipTree(it) }
    }
    includeEmptyDirs = false

    manifest {
        attributes(
            (PluginMetadata.PLUGIN_NAME_KEY)          : "SampleAlgorithms",
            (PluginMetadata.AUTHOR_NAME_KEY)         : "Sample Author",
            (PluginMetadata.PLUGIN_VERSION_KEY)      : "1.0.0 ${getGitHash}",
            (PluginMetadata.ALGORITHM_API_VERSION_KEY): maskingAlgoVer,
            'Build-Timestamp': new java.text.SimpleDateFormat("yyyy-MM-
dd'T'HH:mm:ss.SSSZ").format(new Date()),
            'Created-By'       : "Gradle ${gradle.gradleVersion}",
            'Build-Jdk'        : "${System.properties['java.version']} ($
{System.properties['java.vendor']} ${System.properties['java.vm.version']})",
            'Build-OS'         : "${System.properties['os.name']} $
{System.properties['os.arch']} ${System.properties['os.version']}",
        )
    }
}

dependencies {
    compileOnly ('com.google.code.findbugs:jsr305:3.0.2')
    compileOnly ('com.delphix.masking:masking-algorithm-api:' + maskingAlgoVer)
    compileOnly ('com.fasterxml.jackson.core:jackson-annotations:' + jacksonVer)

    compile 'com.google.guava:guava:' + googleGuavaVer

    testImplementation 'com.google.code.findbugs:jsr305:3.0.2'
    testImplementation 'com.delphix.masking:masking-algorithm-api:' + maskingAlgoVer
    testImplementation 'com.fasterxml.jackson.core:jackson-annotations:' + jacksonVer
    testImplementation 'junit:junit:' + junitVer
    testImplementation "com.google.truth:truth:" + googleTruthVer
}

```

How this works:

- The "from { ... }" property of the **jar** section instructs gradle to include all classes needed at runtime in the plugin JAR file.
- In the **dependencies** section, packages comprising the interface between the plugin and Masking Engine are listed as *compileOnly*. This excludes them from the runtime environment and causes them to be omitted from the plugin JAR file.
- The third-party code dependency on the popular Google Guava library is listed as *compile*, causing it to be included in the plugin JAR file.



Variables defining the package dependency versions are typically read from the *gradle.properties* file.

Plugin Metadata

When a plugin is built for use with the Masking Engine, it is critical that certain metadata be included in the plugin JAR file. This includes certain attributes in the JAR's manifest, as well as the service discovery file used to determine which classes in the JAR are directly usable by the Extensibility Framework.

Manifest Attributes

Java archives carry metadata attributes in the manifest file, located at *META-INF/MANIFEST.MF* in the archive file. Some of these attributes are required or at minimum quite useful in a plugin's manifest.

The following attributes carry special meaning to the extensibility framework when present in a plugin's manifest. Care must be taken to ensure they are set to valid and meaningful values for any plugins intended for production use. Additional attributes may be supported in the future. Any future attributes introduced for anything beyond a purely informational purpose will be of the format "Delphix-*" to avoid conflict with any preexisting usage.

Recognized Manifest Attributes

Attribute	Meaning	Example Value
Delphix-Plugin-Name	The default name of the plugin. This name will be used on the Masking Engine unless overridden at plugin install time. All plugin names beginning with the string "dlpx" are reserved for future use by modules delivered with the Delphix Masking Engine product.	SamplePlugin
Implementation-Vendor	The individual or organization that authored the plugin module.	Sample Inc.
Implementation-Version	The version of the plugin. This is an entirely free-form string, limited to 255 characters.	1.0.0-SNAPSHOT
Delphix-Algorithm-API-Version	The version of the Delphix Masking Plugin API used by the plugin. This value must be present and represent a valid API version.	1.0.0



The `maskScript init` sub-command adds logic to the gradle build files to ensure that the project build inserts the correct attribute values into the plugin manifest.

Versioning

In order to ensure compatibility between each software component in the extensible algorithms system, careful use of versioning must be enforced. The goals are twofold; first, to ensure that the version of each software module is visible, and second, to ensure that incompatible plugins are detected and prevented from running. The software modules particular to extensible algorithms - the Masking Plugin API and Masking Algorithm SDK, use a version number in the following format: Major.Minor.Micro, with an option *-TEXT* notation.

Version information for the plugin, including the plugin version and the version of the Masking Plugin API it was built against, are embedded in the plugin JAR file as [metadata](#).

Table of Versioned Objects

The following table explains how each software module is versioned, and what enforcement takes place:

Software Module	Example Values	Details
Delphix Masking Engine	6.0.3.0	None, however the Masking Algorithm API version is fixed for each particular Delphix Masking Engine release.
Masking Plugin API	1.0.0	The version of the Masking Plugin API used to build a plugin must be embedded in each Plugin. This is done automatically when plugins are built using the Masking Algorithm SDK. Currently, as only one version of the Masking Plugin API exists, the only enforcement in place ensures that a valid version has been embedded in the plugin metadata. In the future, should it be necessary to make backward incompatible changes to the Masking Plugin API, a support matrix will be established and enforced.
Masking Algorithm SDK	1.0.0	Each version of the Masking Algorithm SDK will have a maximum version of the Masking Plugin API it can handle, and will refuse to work with future versions. This is not currently enforced.
Plugins	Author defined	No enforcement. The plugin version will be visible using the Masking API GET operation on the <i>Plugin</i> endpoint.

Ensuring Plugin Compatibility

As the implementation of a plugin evolves, it's natural for the software to change. Changes like bug fixes, performance improvements, etc. can typically be made transparently, without endangering backward compatibility. Challenges can arise when it is necessary to change the configuration schema for a framework within a plugin or remove certain algorithm frameworks or instances from a plugin. This section will detail some best practices for maximizing backward compatibility in each case.

In this case, backward compatibility means that it is possible to upgrade the plugin to a new version on the masking engine without removing the previous version of the plugin. While it is always possible to replace a plugin by removing the old plugin and installing a new one, this requires manual reconstruction of any inventory that references the algorithms based on the old plugin, which can be very labor-intensive.

Schema Changes

A schema change means altering the set of configuration parameters exposed by an algorithm framework. For example, this might be done to add a case-insensitive flag to an algorithm that processes Strings. In order to make this kind of change while preserving backward compatibility, the following rules must be followed:

- Existing configurable variables cannot be removed or modified. These are the class fields with to which the @JsonProperty annotation has been applied.
- New configurable variables may be added, but they must have a default value, so that applying a JSON document lacking a value for the new field results in a valid instance.

If changes must be made that do not meet these requirements, it may be preferable to expose the new or modified functionality as a new algorithm framework, rather than changing the existing one.

Component Removal

Component removal means removing an algorithm framework or a built-in instance provided by an existing framework. When this happens, updating to the new version of the plugin will be blocked if any of the removed objects are in use by the Delphix Masking Engine. This includes references in Inventory, Domains, and any File Formats. In addition, the presence of any user-created algorithms based on a framework which is removed in the new plugin version will block updating.

Plugin Naming

One last compatibility concern is the potential for a plugin name to clash with the name of plugins delivered by the Delphix Masking Engine product. Such a clash would make it impossible to upgrade the engine to a new version without first removing the conflicting user installed plugin. To avoid this concern, avoid embedding any default plugin name beginning with the string "dlpx".

Setting up your development environment

This section describes the step-by-step process for setting up the development environment that was used to develop and test many of the procedures in this Guided Tour. A rich set of tools exist to support Java development, so this is by no means the only development environment possible.



As of version 6.0.3.0, the Delphix Masking Engine's JVM version is 8, so it's important to build a plugin compatible with Java 8.

Downloading and installing tools

- Download and install the Oracle [Java JDK](#) or [OpenJDK](#) from the AdoptOpenJDK Project. Make sure you install **Java 8** version.
- Download and install [IntelliJ IDEA Community Edition](#) for your OS. These instructions are known to work for version 2019.3.

Creating a new project

Identify the root directory of your project code. For example, if you used the instructions to create a new [algorithm](#) or [driver support](#) project, this directory is referred to as **proj_dir**.

Start IntelliJ IDEA. A pop-up should appear.

1. Select **Project Settings > Project > Project SDK > New > JDK**
2. Provide the path to the JDK you installed earlier (e.g. /Library/Java/JavaVirtualMachines/jdk1.8.0_60.jdk/Contents/Home). Select **OK**.
3. For **Project language level**, set to "8 - Lambda, type annotations etc."
4. In the IntelliJ IDE, select **Import Project** from the pop-up.
5. Provide the path to the root of your project, for example, **proj_dir**.
6. Select **Import project from external model** and **Gradle**, and then select **Next**.
7. After the project is imported, a directory tree is shown on the left panel.

At this point, the development environment should be ready to use.

Enabling remote debugging

It is often useful to enable remote debugging, which allows the IDEs debugger to attach to a running **maskApp** or **maskScript** process. To enable remote debugging, certain environment variables must be set. In both cases, the value `5005` can be replaced with the value of any open TCP port.

For **maskApp**

```
MASK_APP_OPTS='-Xdebug -Xnoagent -Djava.compiler=NONE
-Xrunjdp:transport=dt_socket,server=y,suspend=n,address=5005'
```

For **maskScript**

```
MASK_SCRIPT_OPTS='-Xdebug -Xnoagent -Djava.compiler=NONE
-Xrunjdp:transport=dt_socket,server=y,suspend=y,address=5005'
```

 These settings will cause the maskScript to suspend at startup to allow time to attach the debugger.

Algorithms (Authoring extensible plugins)

As of release 6.0.3.0, the Continuous Compliance Engine supports the installation of plugins, written in Java, that provide new masking algorithms. This feature is referred to as Extensible Algorithms. This section of the documentation details all aspects of masking algorithm plugin usage and development. The *Guided Tour* portion of the [workflows section](#) walks the user through the basic process of building a simple plugin and installing it onto the Continuous Compliance Engine. Other sections explore in-depth topics such as making algorithms configurable, consuming input files, etc.

This documentation assumes the reader has some familiarity with Java development as well as operation of the Delphix Masking Engine via both the UI and Web API Client. The reader should also understand the security requirements associated with any new algorithms being developed.

The Extensible Algorithms framework is designed to replace the custom algorithm (aka. mapplets) feature by providing richer functionality, greatly simplifying algorithm development, and ensuring long-term maintainability of plugins. The end-of-support of custom algorithms will occur in release 6.0.15.0 of the Continuous Compliance Engine.

SDK Features

The Masking Algorithm SDK provides a number of useful functions that aid development of new algorithms for the Continuous Compliance Engine. It is available on the Delphix software [download site](#).

- Creation of empty "skeleton" projects, with build files - the `maskScript init` sub-command
- Creation of empty class files for algorithms - the `maskScript generate` sub-command
- Testing of masking algorithms without a masking engine
 - The `maskApp CLI`
 - The `maskScript mask` sub-command
- Uploading of plugins to the masking engine - the `maskScript install` sub-command
- Sample algorithms that illustrate the usage of key features of the Masking Plugin API

Getting more information

Several other sources of information are available to aid in plugin development:

- The `README.md` file under `docs` in the Algorithm SDK download archive
- The [Masking Plugin API Javadoc](#)
- Invoke **maskScript** (located under `sdkTools/bin` in the SDK download) with the `-h` option for usage help
- Type help at the **maskApp** (also under `sdkTools/bin` in the SDK download) command prompt

The MaskingAlgorithm Java Interface

Any Java class that should be recognized as a masking algorithm (whether stand-alone or configurable) must implement the **MaskingAlgorithm** interface. This interface is parameterized with the data type the algorithm masks, which defines the input and output data type of the **mask** method. The full details of this interface are described in the [Masking Plugin API Javadoc](#)

Core Data Types

The Delphix Masking Engine is designed to support a wide and extensible set of data sources, which naturally encode data in a variety of different formats. In order to simplify algorithm development, while maintaining the ability to mask data from many sources, we've identified a core set of data formats which are likely to require different masking treatment and ensured that the Extensible Algorithm framework converts all data to/from these types as needed. These types define the allowed parameterization of the **MaskingAlgorithm** Java interface.

Each masking algorithm class is defined to mask exactly one of the following data types:

- Binary data - **java.nio.ByteBuffer**
- String data - **java.lang.String**
- Numeric data - **java.math.BigDecimal**
- Date time data - **java.time.LocalDateTime**
- Multi-column data - **com.delphix.masking.api.plugin.utils.GenericDataRow** (See Multi-Column Masking section)

Each algorithm is expected to input, process, and emit objects of one of the above Java types, but is free to use any intermediate types as needed to access library methods. Because it is frequently the case that data of one type is stored in databases or documents in a type other than its most natural native type (ex. dates stored in VARCHAR fields, or numbers stored as text in a CSV file), the masking framework that executes these algorithms is capable of performing a number of automatic type conversions, detailed in the next section. This allows algorithms written to process one data type to handle data of other types, with no additional work required of the algorithm author.

Supported Automatic Type Conversions

Algorithm Native Type	Supported Type	Notes
ByteBuffer	String	Algorithm receives the UTF-8 encoded value of the String and is expected to return a valid UTF-8 ByteBuffer.
LocalDateTime	String	The correct date format must be assigned to the field or column in the masking inventory.
LocalDateTime	Compatible numeric types	A compatible date format, such as <i>yyyyMMdd</i> , must be assigned to the column in inventory.
BigDecimal	All numeric types	Upconverted to BigDecimal. Out of range values after masking are truncated to fit the range of the underlying type.
BigDecimal	String	String value is converted to a number.

Special Case Values

In order to allow algorithms to implement special handling for null, empty, and special case values, these values are presented to the masking algorithm unmodified. Algorithms should be prepared to process the full range of input values possible for the input type. In practice, this means that most **mask** method implementations will begin with a null check on the *input* value, prior to attempting to use the input - for example, by calling **input.length()** or similar. It is perfectly acceptable and commonplace to return null in the case where the mask input is null.

Method Overview

This section provides a high-level overview of the methods in the **MaskingAlgorithm** interface. For complete details, consult the Masking Plugin API Javadoc included in the Algorithm SDK archive.

- *getName* and *getDescription* - These methods are used to determine the name and description of frameworks and algorithm instances included in the plugin. For user-created instances, these methods are never called.
- *getDefaultInstances* and *getAllowFurtherInstances* - These methods control the set of instances of the algorithm framework that are defined by the plugin, and whether the user should be allowed to create additional instances.
- *validate* - This method is called after configuration is applied to allow the algorithm class to check whether the injected configuration is valid.
- *setup* and *tearDown* - These methods are called before the algorithm object is used for masking, and after, respectively. Typically, any resources, such as input files, are acquired during *setup* and released during *tearDown*.
- *mask* - This is the method that does the actual data masking in the algorithm class. The input and output values are parameterized for type safety as described above
- *maskBatch* - This method is called to perform masking in situations when it is possible for the caller to build a collection of input values to mask in a single method call. A default implementation is provided that simply calls the *mask* method on each value in the batch.
- *listMultiColumnFields* - This method needs to be implemented only for Multi-Column Algorithms. It returns a list of **AlgorithmLogicalField** objects that define the set of fields that the multi-column algorithm masks.

The following methods are available but deprecated:

- *listMaskedFields* - This method needs to be implemented for Multi-Column Algorithms. It returns a map of field names (`String`) to the Core Data Type. This method does not need to be implemented if not implementing a Multi-Column Algorithm. Implement *listMultiColumnFields* instead.
- *listReadOnlyFields* - Similar to `listMaskedFields` but optional for Multi-Column Algorithms. Fields returned by this method are read-only and cannot be changed. Implement *listMultiColumnFields* instead.

The Life Cycles of Algorithm Objects

The Extensibility framework uses objects classes implementing **MaskingAlgorithm** interface for several distinct purposes. These object life cycles are as follows:

Plugin Discovery

This occurs when the extensibility framework evaluates the capabilities present in a **MaskingAlgorithm** class.

1. Java object creation - an object of the algorithm class is created
2. *getName* - determines framework name
3. *getDescription* - determines framework description
4. *getDefaultInstances*- determines all plugin-provided algorithm instances. For each instance:
 - a. *getName* - determines instance name
 - b. *getDescription* - determines instance description

- c. *validate* - ensure object passes validation
 - d. Serialize configurable fields - these are saved as a JSON document defining the instance's configuration
 - e. Disposal - the Java object is discarded
5. *getAllowFurtherInstances* - determines whether the framework is visible in the *algorithm/framework* API endpoint
 6. Disposal - the Java object is discarded

User Algorithm Creation

This life cycle occurs whenever a user attempts to create a new instance of a plugin algorithm framework. The algorithm definition is saved only if each step succeeds.

1. Java object creation - an object of the algorithm class is created
2. Configuration injection - the values in the user-provided JSON document are injected into the object
3. *validate* - the object's *validate* method is called
4. Disposal - the Java object is discarded

 The *setup* method is not executed when a user-defined instance is created.

Algorithm Use

This is the life cycle of an algorithm object when used to mask data.

1. Java object creation - an object of the algorithm class is created
2. Configuration injection - the saved JSON document defining this instance is injected in the object
3. *setup* - the *setup* method is called once
4. *mask* - the *mask* method is called on each value to be masked
5. *tearDown* - the *tearDown* method is called once
6. Disposal - the Java object is discarded

 It should be noted that a distinct Java object is created for each application of a masking algorithm during Job execution. For algorithms that create or load a large amount of state, this can result in significant memory usage storing redundant data for each instance. This can be avoided using a class level static cache to store data; the instance name, which can be retrieved during *setup* from the **ComponentService** interface object, can be used as an access key for data cached in this way.

Multi-Column Masking

It is possible to write an algorithm that masks data that depends on other column(s) values. In order to account for the different possible data types, we use an object called a `GenericDataRow`.

Generic Data

A `GenericDataRow` is a map of field names (`String`) to `GenericData` objects. Each `GenericData` object contains the value, along with methods to return the respective typed object. When accessing the value from a `GenericDataObject` it will be necessary to read it into a Core Data Type. To do so, use one of the following methods:

- `getStringValue()`
- `getBigDecimalValue()`

- `getLocalDateTimeValue()`
- `getByteBufferValue()`

Once the value has been masked it should be re-set by calling `setValue` and passing as an argument the value as a Core Data Type.

Batch Masking

By overriding the `maskBatch` method in the `MaskingAlgorithm` interface, an algorithm implementation may increase performance or efficiency in cases where the underlying masking operation may be performed more optimally on multiple values per method call. A common example of this is when the algorithm is accessing an external API to perform masking; in this case, masking multiple inputs per method call allows the access latency of the API to be incurred only once for the entire batch of inputs.

The `maskBatch` method is called with a `MaskingBatch` object parameterized by the same Java type used in the `MaskingAlgorithm` interface definition. The `MaskingBatch` object provides the following methods to facilitate masking:

- `size` - returns the size of the batch of values
- `getValue` - returns the value to be masked at a particular index in the batch
- `setValue` - sets the mask result at a particular index in the batch
- `setError` - indicates that an error occurred when masking the input value at a particular index in the batch

The default implementation of `maskBatch` in the `MaskingAlgorithm` interface provides a simple example of how to use these methods.

 The masking engine will not utilize the `maskBatch` method or create a batch with size greater than 1 in all cases. Batch masking is only supported for some job configurations, so it is critical that the `mask` method also be implemented for all algorithms. It is strongly recommended that the `mask` and `maskBatch` method be implemented to produce the same mask results given the same inputs.

Batching is currently supported for these job types:

- All Database masking jobs
- Delimited File masking jobs when no more than one body record type is defined
- Fixed-Width File masking jobs when no more than one body record type is defined

Batch size is equal to the job's `Row Limit` divided by 5, or equal to 2000 when the `Row Limit` is disabled; this is the guaranteed lower bound for batch size, assuming at least that number of inputs are available. Typically, the size of the final batch in a job will be larger.

SDK Workflows (Algorithms)

This section is intended to walk a developer through several workflows using the Delphix Algorithm SDK, such as creating a new algorithm plugin and installing it on a Continuous Compliance Engine. Once an algorithm plugin has been installed, the included algorithms function as expected; they may be assigned to domains and inventory in the normal fashion.

In order to develop and deploy algorithm plugins, you will interact primarily with two tools - the Masking API client, and the Masking Algorithm SDK. The Masking API client is a long-standing feature that allows interactive execution of API operations on the Continuous Compliance Engine, while the Masking Algorithm SDK is a new software package created specifically to aid in algorithm development.

Outline for a guided tour

By following the steps in the outline below, you can tour the basic functionality provided by the Extensible Algorithm feature and Algorithm SDK.

1. Create an algorithm plugin by choosing one of two options:
 - a. [Building the sample algorithm project](#)
 - b. [Creating and building your own algorithm project](#)
2. [Run the algorithm plugin using maskApp](#)
3. [Install the newly created plugin on the Continuous Compliance Engine](#)
4. [View and manage the plugins on a Continuous Compliance Engine using the API Client](#)
5. [Upload multiple plugin in SDK](#)

Building the sample plugin (SDK workflows/Algorithms)

The Algorithm SDK contains a buildable Sample Algorithm Plugin with a number of functional algorithms illustrating the features of the Extensibility Framework. These simple commands build the plugin containing the sample algorithms.

Starting from ***sdk_root***:

```
$ cd samples
$ ./gradlew :algorithm:jar
```

This creates the Sample Algorithm plugin JAR file ***sdk_root***/samples/build/libs/algorithm.jar.

The Sample Algorithm project provides a convenient way to see a working example plugin.

i While it is possible to modify these algorithms by changing the Java source and rebuilding the plugin, when starting a new project to develop one or more standalone algorithms, it is highly recommended that you [create your own project](#) rather than modifying files in the Sample Algorithm project subtree. This will prevent the loss of customizations to the project build files should you chose to install a new version of Masking Algorithm SDK over your existing SDK directory.

Creating a New Project (SDK workflows/Algorithms)

This section describes how to create a brand new Java project for a new masking algorithm plugin. We will use the `maskScript` utility to create a skeleton project and an empty algorithm class in that project.

Creating the Project

Before you begin, you'll want to pick a name for your project, and an **empty** directory (outside of the Masking SDK source tree) where your project will be created. Once you've done this, run this **maskScript** command:

```
$ maskScript init -t algorithm -d <project path> -n <project name> -a <author name>
-v <version>
```

For example, this command will create a project named `demoProject` in the `demo-proj` subdirectory of your home directory.

```
$ maskScript init -d $HOME/demo-proj -n demoProject -a "Demo Author" -v 1.0.0
```

For the rest of this section, we'll assume a new project has been created under **proj_dir**. Change your working directory to **proj_dir**. You'll notice that the project is created with a sample algorithm file `proj_dir/src/main/java/com/sample/SampleAlgorithm.java`. It's possible to build this into a usable plugin by running:

```
$ cd <proj_dir>
$ ./gradlew jar
```

But let's create our own, brand new algorithm.

Creating an Algorithm Class

For this part of the tour, we're going to create a new algorithm named Clobber. First, we'll run the `maskScript` utility to create a skeleton class file:

```
$ cd <proj_dir>
$ maskScript generate -p com.delphix.demo -c Clobber -v String -s .
```

By convention, the class file `Clobber.java` will be created under a sub-directory path based on the package name, so it might be helpful to use the `find` command to locate it:

```
$ find . -name Clobber.java
./src/main/java/com/delphix/demo/Clobber.java
```

The initial content of this file is:

```
$ cat ./src/main/java/com/delphix/demo/Clobber.java

package com.delphix.demo;

import com.delphix.masking.api.plugin.MaskingAlgorithm;
```

```

import java.lang.String;
import javax.annotation.Nullable;

public class Clobber implements MaskingAlgorithm<String> {

    /**
     * Masks String object
     * @param input The String object to be masked. This method should handle null
    inputs.
     * @return Returns the masked value.
     */
    @Override
    public String mask(@Nullable String input) {
        // TODO: change the default implementation.
        return input;
    }

    /**
     * Get the recommended name of this Algorithm.
     * @return The name of this algorithm
     */
    @Override
    public String getName() {
        // TODO: Change this if you'd like to name your algorithm differently from the
    Java class.
        return "Clobber";
    }
}

```

Customizing the Algorithm Class

The first thing to notice about the skeleton algorithm is that the `mask` method just returns the input. This means no masking will be done, so this will certainly need to change. We're going to create an algorithm that overwrites the entire input `String` with the first letter of that `String`. This replaces the skeleton `mask` method with:

```

@Override
    public String mask(@Nullable String input) {
        // Always be ready to handle null or empty input
        if (input == null || input.length() < 2) {
            return input;
        }

        char firstChar = input.charAt(0);
        StringBuilder result = new StringBuilder();

        for (int i = 0; i < input.length(); i++) {
            result.append(firstChar);
        }

        return result.toString();
    }

```

The algorithm name "Clobber" is fine, so we can just delete the TODO comment in the getName() method.

Now, we'll rebuild the project to include this new algorithm in the plugin JAR:

```
$ ./gradlew jar
```

This creates or updates the plugin JAR file **proj_dir**/build/libs/demoProject.jar

Service discovery (SDK workflows/Algorithms)

Java service discovery is used to determine which classes in the plugin JAR present relevant functionality to the Delphix Masking Engine. When a plugin is loaded, the file *com.delphix.masking.api.plugin.MaskingComponent* under *META-INF/services* in the JAR is consulted for a list of classes that implement the **MaskingComponent** interface. As **MaskingAlgorithm** includes this interface, each algorithm in the plugin will be discovered this way. In the future, this mechanism may be expanded to support additional types of components beyond algorithms.

- When the maskScript *generate* sub-command is used to create a new algorithm class, the service discovery metadata file is automatically updated.

If an algorithm class is missing from the services file, it will not be usable when the plugin is loaded. It is essentially invisible to the extensibility framework. If a class is mentioned in this file but not present in the JAR, the plugin will fail to load. There is a fallback during plugin loading that will scan the entire JAR for algorithms if the services file is not present. This fallback may be removed in the future and should not be relied on.

Running an Algorithm using the SDK tools (SDK workflows/Algorithms)

It will often be more convenient to use the SDK utilities to test an algorithm since this avoids the need to install or update your plugin, create masking inventory, and execute jobs on the Continuous Compliance Engine. This can be done interactively using `maskApp`, or entirely from the command line using `maskScript`.

Using `maskApp` to Test an Algorithm

The **maskApp** application is interactive and may be launched with no parameters from the shell:

```
$ maskApp
```

After a moment, the application will print a banner and prompt for input. Currently, the only sub-command supported is `mask`. In order to use this command, you'll need to know the location of the plugin JAR file on the filesystem.

```
MASKING-APP:> mask -j <plugin_jar_location>
```

You will be prompted to select an algorithm. This example runs `maskApp` in **proj_dir** and uses the JAR file created with the [Create a new project](#) workflow:

```
MASKING-APP:> mask -j build/libs/demoProject.jar
/Users/*****/demo-proj/build/libs/demoProject.jar
Loaded plugin demoProject version 1.0.0 (API version: 1.0.0) from [/Users/*****/
demo-proj/build/libs/demoProject.jar]
13:17:00.707 [main] INFO global - Loaded plugin demoProject: Plugin {'embeddedName':
'demoProject', 'version': '1.0.0', 'author': 'Delphix Dev', 'apiVersion': '1.0.0'}
Framework:
* [0] Clobber
  [1] SampleAlgorithm
Select an algorithm framework: 0
Instance:
* [0] demoProject:Clobber
Select an instance of algorithm framework: Clobber: 0
Selected algorithm: com.delphix.demo.Clobber(Clobber) instance: demoProject:Clobber,
data type: STRING
Input value to be masked('null' for null, 'doneMasking' to finish): Test1
Masked value: TTTT
Input value to be masked('null' for null, 'doneMasking' to finish): test2
Masked value: tttt
Input value to be masked('null' for null, 'doneMasking' to finish): 1 more test
Masked value: 1111111111
Input value to be masked('null' for null, 'doneMasking' to finish):
```

When you invoke the `mask` sub-command, you will first be presented with a list of possible frameworks (aka. Algorithm Components) to choose from. These correspond with the Java classes that implement the **MaskingAlgorithm** interface in the plugin file. Once you have selected a framework, you will be presented with a list of each pre-defined instance to choose from. If the algorithm supports creating new instances, that option will be present as well. Once an instance is selected, you're ready to enter test values and see the masked result.

- ☐ In order to be usable, each class that implements **MaskingComponent** must also be listed in the appropriate service description file. Refer to [this section](#) for details.

The selection marked with a star is the default selection; you may always press return at the prompt to make the default selection.

- ☐ When a Multi-Column algorithm is selected, the prompt will contain the order to provide the input values in. They should be placed on a single row, separated by a comma.

```
Enter CSV-formatted values for the following columns in the following
order: date1(LOCAL_DATE_TIME), date2(LOCAL_DATE_TIME)
```

- ☐ Missing Algorithms

If an algorithm seems to be missing from the list, or an algorithm's behavior does not seem to match the latest version of the code, it may be necessary to rebuild the plugin JAR:

```
$ cd ; ./gradlew="">>;>
```

Running an Algorithm using maskScript

Algorithms may also be tested using the SDK **maskScript**. The **maskScript** utility is non-interactive, which lets you conveniently process input files using a masking algorithm. Each input line is considered a separate value to be masked. The algorithm framework and instance are selected using command-line options. This example uses the "Redaction X" algorithm instance from the Sample Algorithm plugin. This plugin can be built using the process described [here](#).

Create a small sample input file:

```
$ echo "Adam
Amy
Brandon" > test_input.txt
```

Mask each line of the file to standard output:

```
$ cat test_input.txt | maskScript mask -j algorithm/build/libs/algorithm.jar -n
"Sample Plugin:Redaction Z"
Jun 19, 2020 1:51:54 PM com.delphix.masking.api.provider.LogService info
INFO: Loaded plugin Sample Plugin: Plugin {'embeddedName': 'Sample Plugin', 'version':
'1.0.0', 'author': 'Delphix', 'apiVersion': '1.0.0'}
ZZZZ
ZZZ
ZZZZZZZ
Jun 19, 2020 1:51:54 PM com.delphix.masking.api.provider.LogService info
INFO: StringRedaction: Masked a total of 3 values
```

- When masking using a Multi-Column algorithm, the inputs in the file must be provided in the same format they would be provided when using the `mask` subcommand of the `maskApp` (i.e.: comma-separated list of expected values). If unsure of the order, use the `mask` subcommand in the `maskApp` to see what the expected order is.

i redirecting Information Messages

To remove all informational message from the output, redirect standard error to `/dev/null` or an alternate location:

```
$ cat test_input.txt | maskScript mask -j algorithm/build/libs/  
algorithm.jar -n "Sample Plugin:Redaction Z" 2>/dev/null
```

Installing multiple plugins onto the Delphix Masking engine (SDK workflows/Algorithms)

Starting SDK version 1.4.0 (corresponding to the Masking Engine version 6.0.8.0) Delphix has implemented multiple plugins upload within SDK, where the Delphix provided `dlpx-core` plugin is uploaded by default. That gives an option to chain multiple extensible algorithms, even if those are based different plugins.

Load multiple algorithm plugins

1. Using `maskApp` to Test an Algorithm The **maskApp** application is interactive and may be launched with no parameters from the shell: `$ maskApp` After a moment, the application will print a banner and prompt for input. Currently, the only sub-command supported is `mask`. In order to use this command, you'll need to know the location of the plugin JAR file on the filesystem.
2. Upload a desired algorithm plugin `bash MASKING-APP:> mask -j <>` This example runs `maskApp` in **proj_dir** and uses the JAR file created with the [Create a new project](#) workflow:

```
MASKING-APP:> mask -j ./algorithm/build/libs/plugin1.jar
/Users/testuser/delphix/masking-algorithm-sdk/./algorithm/build/libs/plugin1.jar
Loading security manager
Loaded plugin dlpx-core version 1.4.0 (API version: 1.4.0) from [/Users/testuser/
delphix/masking-algorithm-sdk/./sdkTools/build/install/maskApp/lib/delphix-algorithm-
plugin-1.4.0.jar, /Users/testuser/delphix/masking-algorithm-sdk/./algorithm/build/
libs/plugin1.jar]
Loaded plugin plugin1 version 1.4.0 (API version: 1.4.0) from [/Users/testuser/
delphix/masking-algorithm-sdk/./sdkTools/build/install/maskApp/lib/delphix-algorithm-
plugin-1.4.0.jar, /Users/testuser/delphix/masking-algorithm-sdk/./algorithm/build/
libs/plugin1.jar]
Framework:
* [0] dlpx-core:CM Numeric
  [1] dlpx-core:Character Mapping
  [2] dlpx-core:Date Replacement
  [3] dlpx-core:Date Shift
  [4] dlpx-core:Date Shift Discrete
  [5] dlpx-core:Date Shift Variable
  [6] dlpx-core:Dependent Date Shift
  [7] dlpx-core:FullName
  [8] dlpx-core:Name
  [9] dlpx-core:Payment Card
 [10] dlpx-core:Regex Decompose
 [11] dlpx-core:Secure Lookup
 [12] plugin1:Byte Array Redaction
 [13] plugin1:Date Redaction
 [14] plugin1:Number Redaction
 [15] plugin1:Randomized Masking
 [16] plugin1:StringRedaction
Select an algorithm framework:
```

- Algorithm framework name is now prefixed with the plugin name (that framework is originated from). There are always `dlpx-core` plugin frameworks present, since those are provided by default by the Masking Engine. Previously one could only chain the algorithm with the other algorithm provided by the same plugin. Now it is possible to chain plugin's algorithm with the algorithm instance(s), based on the default `dlpx-core` plugin.

It is also possible to upload another plugin along with the already loaded ones: instead of selecting an algorithm framework from the above menu - step back by pressing `Ctrl-C`. That will bring you back to the `MASKING-APP:>` menu (with the `UserInterruptException` notification). That error message will be taken care of in a future releases, providing better way for this step back.

```
org.jline.reader.UserInterruptException
Details of the error have been omitted. You can use the stacktrace command to print
the full stacktrace.
MASKING-APP:>
```

Here it is possible to use similar `mask -j <>` command to upload another plugin:

```
MASKING-APP:> mask -j ./algorithm/build/libs/plugin2.jar
/Users/testuser/delphix/masking-algorithm-sdk/./algorithm/build/libs/plugin2.jar
Loading security manager
Loaded plugin dlpx-core version 1.4.0 (API version: 1.4.0) from [/Users/testuser/
delphix/masking-algorithm-sdk/./sdkTools/build/install/maskApp/lib/delphix-algorithm-
plugin-1.4.0.jar, /Users/testuser/delphix/masking-algorithm-sdk/./algorithm/build/
libs/plugin2.jar]
Loaded plugin plugin2 version 1.4.0 (API version: 1.4.0) from [/Users/testuser/
delphix/masking-algorithm-sdk/./sdkTools/build/install/maskApp/lib/delphix-algorithm-
plugin-1.4.0.jar, /Users/testuser/delphix/masking-algorithm-sdk/./algorithm/build/
libs/plugin2.jar]
21:17:37.426 [main] INFO global - Loaded plugin plugin2: Plugin {'embeddedName':
'plugin2', 'version': '1.4.0', 'author': 'Sample Plugin Author', 'apiVersion':
'1.4.0'}
Framework:
* [0] dlpx-core:CM Numeric
  [1] dlpx-core:Character Mapping
  [2] dlpx-core:Date Replacement
  [3] dlpx-core:Date Shift
  [4] dlpx-core:Date Shift Discrete
  [5] dlpx-core:Date Shift Variable
  [6] dlpx-core:Dependent Date Shift
  [7] dlpx-core:FullName
  [8] dlpx-core:Name
  [9] dlpx-core:Payment Card
 [10] dlpx-core:Regex Decompose
 [11] dlpx-core:Secure Lookup
 [12] plugin1:Byte Array Redaction
 [13] plugin1:Date Redaction
 [14] plugin1:Number Redaction
 [15] plugin1:Randomized Masking
```

```
[16] plugin1:StringRedaction
[17] plugin2:MultiColumnDateAlgorithm
[18] plugin2:Numeric Mapping
[19] plugin2:RedactionDB
[20] plugin2:RedactionFile
[21] plugin2:StringHashedLookup
Select an algorithm framework:
```

Example above shows 3 plugins uploaded:

- dlp-core (default)
- plugin1 (uploaded by customer)
- plugin2 (uploaded by customer)

Now it is possible choosing any of those plugins frameworks and their related algorithm instances, as well chaining of those algorithms instances to any configurable extensible algorithm (within the loaded plugins). That behavior simulates Masking Engine where multiple plugins are uploaded.

The technique of algorithms chaining is out of scope of the current description. It's the same as chaining the algorithms belonging to the same plugin. Please refer to the [Algorithm chaining page](#) for chaining details and examples.

Retrieving information about installed plugins (SDK workflows/Algorithms)

The GET endpoints are useful for getting information about plugins. After following the steps in [this section](#) to install the Sample Algorithm plugin, the GET operation will return (elided for brevity):

```
{
  "pluginId": 7,
  "pluginName": "Delphix Sample",
  "originalFileName": "algorithm.jar",
  "originalFileChecksum":
"74df61f436aceb80107c22964c027d32a565d0100de36c7fa42f528327cf2e2a",
  "installDate": "2020-06-19T20:15:58.239+0000",
  "installUser": 5,
  "builtIn": false,
  "pluginVersion": "1.0.0",
  "pluginObjects": [
    {
      "objectIdentifier": "2",
      "objectName": "StringRedaction",
      "objectType": "ALGORITHM_FRAMEWORK"
    },
    {
      "objectIdentifier": "3",
      "objectName": "RedactionFile",
      "objectType": "ALGORITHM_FRAMEWORK"
    },
    {
      "objectIdentifier": "5",
      "objectName": "StringHashedLookup",
      "objectType": "ALGORITHM_FRAMEWORK"
    },
    {
      "objectIdentifier": "7",
      "objectName": "Randomized Masking",
      "objectType": "ALGORITHM_FRAMEWORK"
    },
    {
      "objectIdentifier": "Delphix Sample:Byte Array Redaction",
      "objectName": "Delphix Sample:Byte Array Redaction",
      "objectType": "ALGORITHM"
    },
    {
      "objectIdentifier": "Delphix Sample:Date Redaction",
      "objectName": "Delphix Sample:Date Redaction",
      "objectType": "ALGORITHM"
    },
    {
      "objectIdentifier": "Delphix Sample:Number Redaction",
      "objectName": "Delphix Sample:Number Redaction",
      "objectType": "ALGORITHM"
    }
  ],
}
```

```
{
  "objectIdentifier": "Delphix Sample:Numeric Mapping",
  "objectName": "Delphix Sample:Numeric Mapping",
  "objectType": "ALGORITHM"
},
...
]
```

For each plugin, the plugin metadata, including `pluginId`, `pluginName` and `originalFileChecksum` are displayed first. This is followed by a list of algorithm frameworks included in the plugin, then a list of algorithm instances included in the plugin. The list of frameworks will contain only those frameworks that support the creation of additional algorithm instances as described in [this section](#).

Configurability

Introduction

The Extensible Algorithms feature supports the creation of algorithm frameworks. When an algorithm class is constructed to be a framework, the Continuous Compliance Engine operator may create additional instances of the algorithm - with different configurations - after the plugin has been installed. New instances are created by supplying a JSON document describing a new instance using the POST method of the Algorithm endpoint in the Masking Web API. This may be done using the Masking API client. The JSON schema for configuration is determined by which data members in the framework class are marked as configurable, and may vary from framework to framework.

New algorithms created using the SDK skeleton generator are not, by default, configurable using this mechanism. It is necessary to modify the default implementation of the *allowFurtherInstances* method and mark one or more public data members as configurable. This is described in detail in [the next section](#).

This part of the documentation illustrates what options are available when creating an algorithm to define whether and what kind of configuration is required, and what, if any, default instances should be created. It will also describe how to create instances of a plugin provided algorithm framework using the Masking API Client.

Making an Algorithm Configurable

As described in the introduction, there are a couple of requirements for making an algorithm configurable, so that it will appear as a framework on the Continuous Compliance Engine:

1. The **getAllowFurtherInstances** method in the algorithm Class must return *true*.
2. One or more data members in the algorithm class must be marked *public*, and must be annotated with the **@JsonProperty** (specifically *com.fasterxml.jackson.annotation.JsonProperty*) annotation.

In order to assure that JSON document and schema interpretation is consistent, most JSON handling is done by the Masking Plugin API implementation, rather than the plugins themselves. For each configurable algorithm, the SDK or Continuous Compliance Engine will examine the annotations in the class to determine which values are configurable. Whenever a new instance is created, an attempt is made to apply the user-supplied JSON to the object of the framework class. This includes *some* validation that the supplied JSON matches the expected schema implied by the set of fields marked configurable, however there are some limitations to this validation, as described below.

Example Configurable Algorithm Explained

The concept of configurability can be illustrated using one of the sample algorithms from the SDK as an example - StringRedaction.java in this case:

```
package sample.masking.algorithm.redaction;
...

public class StringRedaction implements MaskingAlgorithm<String> {
    private String name = "StringRedaction";

    @JsonProperty(value = "redactionCharacter", required = true)
    public String redactionCharacter = "specified";

    @Override
    public String getName() {
        return name;
    }

    @Override
    public Collection<MaskingComponent> getDefaultInstances() {
        StringRedaction instanceX = new StringRedaction();
        instanceX.name = "Redaction X";
        instanceX.redactionCharacter = "X";
        return Arrays.asList(instanceX);
    }

    @Override
    public boolean getAllowFurtherInstances() {
        return true;
    }

    @Override
    public String getDescription() {
        return String.format(
```

```

        "Redact String by overwriting with '%s' character",
redactionCharacter);
    }

    @Override
    public String mask(@Nullable String input) throws MaskingException {
        if (input == null) {
            return null;
        }
        StringBuilder returnVal = new StringBuilder();

        for (int i = 0; i < input.length(); i++) {
            returnVal.append(redactionCharacter);
        }
        return returnVal.toString();
    }

    @Override
    public void validate() throws ComponentConfigurationException {
        if (redactionCharacter == null || redactionCharacter.length() != 1) {
            throw new ComponentConfigurationException(
                "redactionCharacter must be a single character");
        }
    }
}

```

This algorithm does simple redaction of the input String, but the redaction character may be configured by creating additional instances with custom values. How this works:

- The Class has a public field `redactionCharacter` annotated with `@JsonProperty`. A default value has been provided so that the **getDescription** method will return a suitable description in both the framework and instance cases.
- The Class's **getDefaultInstances** method defines a single instance, with redaction characters 'X'. This is accomplished by simply returning a list of correctly configured objects. The API framework extracts the object configuration as JSON, and store it for use whenever an instance of "Redaction X" algorithm is needed.
- The Class's **getAllowFurtherInstances** method returns true, making it possible to create additional instances of this algorithm after the plugin is loaded on the Masking Engine using the Masking API via the API client.
- The Class implements a **validate** method to ensure that the supplied configuration value is usable. In this case, the length of the `redactionCharacter` String is restricted to a single character.

Frameworks, Instances, and Configuration Injection

When used as a framework, the algorithm class is instantiated and used without any configuration injection. In the example above, that means that the **getDescription** method will return "Redact String by overwriting with 'specified' character" when the algorithm framework is evaluated. Similarly, **getName** will return "StringRedaction", the name of the framework.

When a runnable algorithm instance is needed, the algorithm class is instantiated, and all saved configuration is injected before any methods are called. This configuration is gathered in one of two ways:

- For statically provided instances embedded in the plugin, the configurable fields of each object returned by the **getDefaultInstances** method are serialized to JSON and saved. Again, only the values of public fields marked with the `@JsonProperty` annotation are extracted this way.

- When the user creates a new algorithm instance using the Masking Web API, the contents of the `algorithmExtension` field of the POST or PUT request is validated and saved for future injection whenever that particular algorithm instance is needed in the future.

Using the above example again, when algorithm instance "Redaction X" is created, the saved values will be injected, so `redactionCharacter` will have the value 'X'.

Validation of Configuration Values

For what the author can only presume to be performance considerations, the major JSON handling libraries perform only minimal validation when objects are deserialized. The practical effect of this is that several aspects of the `@JsonProperty` annotation are not enforced. For example, a property might be marked as required, but an object will be successfully deserialized even when that property is missing from the input JSON. While libraries are available that would allow us to expand the degree to which JSON is validated by the framework, this would make defining the exact set of validations done by the API framework vs. what must be validated in the component's `validate` method even more complex. For these reasons, only minimal input validation is performed by the framework. Plugin authors should validate all aspects of the object's configuration, especially the presence (that is, non-null, non-empty value) of required fields, in the `validate` method implementation.

However, this is not to say that the unenforced properties of the `@JsonProperty` annotation should be omitted. These values are visible in the auto-generated schema for each framework, which is visible using the SDK's `maskApp`, as well as the algorithm/framework endpoint in the Masking API Client, and may be useful for UI generation in the future.

Default Interface Implementations

The Masking Plugin API defines default implementations of **`getDefaultInstances`** and **`getAllowFurtherInstances`** as follows:

```
default Collection<ComponentInstanceDescription> getDefaultInstances() {
    return Collections.singletonList(this);
}

default boolean getAllowFurtherInstances() {
    return getDefaultInstances() == null || getDefaultInstances().isEmpty();
}
```

This means that if neither of these methods is overridden by the masking algorithm class, a single instance capturing whatever default values exist for configurable fields is created by default.

Only algorithms classes that define **`getAllowFurtherInstances`** to return `true` appear as Algorithm Frameworks on the Masking Engine.

Build Dependencies for Configurable Algorithms

When the `maskScript init` sub-command is used to create a new project, the initial build files will may not include the dependencies required for the Jackson `@JsonProperty` annotation. This can be corrected by adding this line to **`proj_root`**/`gradle.properties`:

```
jacksonVer=2.9.5
```

And this line to the **`dependencies`** section at the end of `proj_root`*/`build.gradle`:

```
compileOnly ('com.fasterxml.jackson.core:jackson-annotations:' + jacksonVer)
```

The set of Jackson annotations tested and supported for use in algorithm plugin classes are:

- *@JsonProperty*
- *@JsonPropertyDescription*
- *@JsonFormat* (Useful in specifying formats for Date fields)

Using an Algorithm Framework

When a plugin algorithm supports configuration, it is possible to create new instances of the algorithm on the Continuous Compliance Engine by specifying the desired configuration. This is done using the engine's [Masking Web API](#). Configurable algorithms may also be tested using the **maskApp** and **maskScript** utilities, by providing the desired configuration in an input file or at the command line.

Creating New Algorithm Instances Using the maskApp SDK Utility

When the maskApp utility's `mask` command is invoked and a configurable algorithm is selected, the option will be presented to create a new algorithm instance. This is done by choosing "::Create New Instance:". The algorithm's configuration schema is displayed, and then a valid JSON input must be provided to create the new instances. Rather than entering the literal JSON, the '@' symbol may be used to load the JSON from a file (**@file-path**).

What follows is an example of loading the Sample Algorithm plugin, creating a new instance of the *StringRedaction* framework and masking test values with the new algorithm instance.

```
$ maskApp
... Startup Messages ...
MASKING-APP:> mask -j algorithm/build/libs/algorithm.jar
/Users/jleser/ws/algorithm-sdk/algorithm/build/libs/algorithm.jar
Loaded plugin Delphix Sample version 1.0.0 dea904c (API version: 1.0.0) from [/Users/
jleser/ws/algorithm-sdk/algorithm/build/libs/algorithm.jar]
16:44:47.743 [main] INFO global - Loaded plugin Delphix Sample: Plugin
{'embeddedName': 'Delphix Sample', 'version': '1.0.0 dea904c', 'author': 'Delphix',
'apiVersion': '1.0.0'}
Framework:
* [0] Byte Array Redaction
  [1] Date Redaction
  [2] Number Redaction
  [3] Numeric Mapping
  [4] Randomized Masking
  [5] RedactionFile
  [6] StringHashedLookup
  [7] StringRedaction
Select an algorithm framework: 7
Instance:
* [0] Delphix Sample:Redaction X
  [1] Delphix Sample:Redaction Y
  [2] Delphix Sample:Redaction Z
  [3] ::Create New Instance::
Select an instance of algorithm framework: StringRedaction: 3
The JSON schema of the selected framework is:
{
  "type" : "object",
  "id" : "urn:jsonschema:sample:masking:algorithm:redaction:StringRedaction",
  "properties" : {
    "redactionCharacter" : {
      "type" : "string",
      "required" : true
    }
  }
}
```

```

}
}
Enter config(Prefix with '@' for file location)(Blank for no config): {
"redactionCharacter" : "+" }
Enter instance name: RedactPlus
Algorithm Configuration: {"redactionCharacter":"+"}
Selected algorithm:
sample.masking.algorithm.redaction.StringRedaction(StringRedaction) instance:
RedactPlus, data type: STRING
Input value to be masked('null' for null, 'doneMasking' to finish): Test
Masked value: ++++
Input value to be masked('null' for null, 'doneMasking' to finish): One
Masked value: +++
Input value to be masked('null' for null, 'doneMasking' to finish): TwoThree
Masked value: +++++++

```

Creating New Algorithm Instances on the Continuous Compliance Engine

New instances of plugin frameworks may be created using the Continuous Compliance Engine's Web API's *algorithm* endpoint. This is similar to creating any other algorithm using the *algorithm* API endpoint and may be performed using the API client. Unlike when an algorithm is created using older, built-in frameworks like Secure Lookup:

- The value for *algorithmType* in the JSON request is always "COMPONENT". This is now the default value, so this field may be omitted.
- A value for the field *frameworkId* must be included - this is the integer ID of the framework as provided in the plugin description retrievable using the GET operation on the plugin endpoint, or GET on the *algorithm/frameworks* endpoint.
- The *algorithmExtension* field's contents are used directly as the JSON configuration for the algorithm instance. Unlike other algorithm types, this field does not have a fixed schema for COMPONENT type algorithms. The required schema may be retrieved using the [procedure described below](#).

This example API request, POSTed to the algorithm endpoint, creates a new instance of the StringRedaction algorithm (described above), named "RedactStar" using "*" as the redaction character. In this case, the sample algorithm plugin JAR has already been uploaded, and the StringRedaction framework has id 19:

```

{
  "algorithmName": "RedactStar",
  "algorithmType": "COMPONENT",
  "description": "Redact with the star character",
  "frameworkId" : 19,
  "algorithmExtension" : {
    "redactionCharacter": "*"
  }
}

```

Discovering the algorithm extension API Field Schema

The Masking Web API *algorithm/framework* endpoint has the ability to show the JSON Schema for each algorithm framework implemented using the extensibility mechanism. By default, the schema is not included, but by setting *include_schema* true, the schema may be retrieved. Here is the GET API result, including schema, for the StringRedaction framework used above:

```
{
  "frameworkId": 19,
  "frameworkName": "StringRedaction",
  "frameworkType": "STRING",
  "plugin": {
    "pluginId": 47,
    "pluginName": "algorithm"
  },
  "extensionSchema": {
    "id": "urn:jsonschema:sample:masking:algorithm:redaction:StringRedaction",
    "properties": {
      "redactionCharacter": {
        "type": "string",
        "required": true
      }
    }
  }
}
```

This schema is generated automatically using the annotated public fields in the framework class.

Using Multi-Column Algorithms

To be able to configure and use the Multi-Column (MC) Algorithms one should be familiar with the following themes:

- Extensible Algorithms in general
- Their creation using the [Masking SDK](#)
- Extensible Algorithms [Plugin installation](#)
- [Masking API Client](#) (optional)

Logical Fields

A sample instance (serving as an example) of the MC algorithms is in the Masking SDK distribution, named "MultiColumnDateAlgorithm". That framework (the instance is based on) defines two fields:

```
@Override
public List<AlgorithmLogicalField> listMultiColumnFields() {
    /*
     * Here we define the column names to be used in the algorithm. These names
     * are only used to reference the
     * columns within the algorithm and do not need to correspond to the names
     * of the columns on the data source.
     * For example, our data source may call these 2 fields "dateOfBirth" and
     * "dateOfDeath", however within the
     * algorithm implementation they will be referenced as "startDate" and
     * "endDate" (see mask method to see how
     * this is used).
     */
    return ImmutableList.of(
        new AlgorithmLogicalField("startDate", MaskingType.LOCAL_DATE_TIME),
        new AlgorithmLogicalField("endDate", MaskingType.LOCAL_DATE_TIME));
}
```

In that example, the fields "startDate" and "endDate" are logical fields, defined by the framework. If one doesn't have access to the source code of the framework, it's possible to find the logical field names (and their types) using the *Masking API*: `GET /algorithms/{algorithmName}` endpoint.

The API provides a five-argument constructor for **AlgorithmLogicalField** that allows for fields to be marked as *read-only* and/or *optional*, as well as to provide a short documentation string for the field's usage. The Extensibility SDK provides an example algorithm that demonstrates this called *MultiColumnRedaction.java*.

Let's suppose you already have an instance of multi-column Algorithm installed. That might happen in any of the following two cases:

- The Plugin you've installed contains a default instance for MC algorithms.
- The Plugin you've installed contains only a framework for configurable MC algorithms. In that case, you've configured an instance of the algorithm.

Let's take as an example "MultiColumnDateAlgorithm" algorithm mentioned above (plugin is named "sample" in that example). Retrieving its info using the `GET /algorithms/{algorithmName}` endpoint returns:

```
{
  "algorithmName": "Sample Plugin:MultiColumnDateAlgorithm",
```

```

"algorithmType": "COMPONENT",
"isTokenizationSupported": false,
"pluginId": 11,
"fields": [
  {
    "fieldId": 5,
    "name": "startDate",
    "type": "LOCAL_DATE_TIME",
    "isReadOnly": false,
    "isOptional": false
  },
  {
    "fieldId": 6,
    "name": "endDate",
    "type": "LOCAL_DATE_TIME",
    "isReadOnly": false,
    "isOptional": false
  }
],
"algorithmExtension": {}
}

```

Here we can see the information structure for the logical fields, defined by the current framework. We will use that data when configuring the Inventory fields.

i Previous versions of the Extensibility API required two methods - *listMaskedFields* and *listReadOnlyFields* - to be implemented when creating a multi-column algorithm. These methods are now deprecated, and *listMultiColumnFields* is preferred way for multi-column algorithms to define their fields. However, existing algorithms that use the old methods should continue to function normally.

Configuring column metadata for MC algorithm

To configure the involved column (i.e. masked and read-only columns) - we should update the column's metadata with the following information:

```

"algorithmFieldId"
"algorithmGroupNo"
"algorithmName"
"domainName"

```

The last two fields are the regular configuring fields for masked columns. Let's look closer to the newly introduced fields for MC:

- `algorithmFieldId` is a fieldId for the corresponding logical field. For example for "startDate" from the example above its value is 5.
- `algorithmGroupNo` is a group number (integer) for the columns treated by the same algorithm instance. It is introduced for cases where we might have multiple columns of a similar type, which are masked by the different Masking Jobs using the same algorithm. In such a case that's important to unite the columns per algorithm run, by assigning the same group number.

There are two supported methods to configure the columnMetadata for the masked table inventory:

- Via API

- Via UI

Configuring columnMetadata for MC algorithms via API

Below is the example of the column metadata before it's configured for MC algorithm:

```
Response Body

{
  "columnMetadataId": 63,
  "columnName": "DATA00",
  "tableMetadataId": 19,
  "dataType": "VARCHAR2",
  "columnLength": 100,
  "isMasked": false,
  "isProfilerWritable": true,
  "isPrimaryKey": false,
  "isIndex": false,
  "isForeignKey": false
},
```

Let's associate that field with the logical field `startDate` (fieldId=5) from the snapshot above, by adding the mentioned fields:

Response Body

```

{
  "columnMetadataId": 63,
  "columnName": "DATA00",
  "tableMetadataId": 19,
  "algorithmName": "sample:MultiColumnDateAlgorithm",
  "algorithmFieldId": 5,
  "algorithmGroupNo": 1,
  "domainName": "DOB",
  "dataType": "VARCHAR2",
  "dateFormat": "yyyy-MM-dd",
  "columnLength": 100,
  "isMasked": true,
  "isProfilerWritable": true,
  "isPrimaryKey": false,
  "isIndex": false,
  "isForeignKey": false,
  "domainAssignedBy": "admin_user"
},
    
```

i For the masked column, the *isMasked* field should be manually changed to *true*, while for read-only field it stays *false*.

If at this point an inventory for the masked table is checked in the UI - the configured (via API) inventory will be displayed there:

Home > Environments > test1 > Inventory > test

test

Filter By: All Fields Masked Fields Auto User

Column	Data Type	Algorithm	Edit
DATA00	VARCHAR2 (100)	sample:MultiCo...nDateAlgorit...	
DATA01	VARCHAR2 (100)		
ID	NUMBER (38)		

Select Rule Set: test

Filter Contents: Search By Name, Search Alphabetically

Configuring columnMetadata for MC algorithms via UI

The same columnMetadata configuration can also be made via the UI. As with other algorithms one has to choose the Domain and Algorithm values, applied to the current column. If a Multi-Column algorithm has been chosen, the following additional two fields will need to be filled out:

- **Select Logical Field** dropdown, where the corresponding logical field to be selected.
- **Algorithm Group** window, where algorithmGroupNo value to be entered.

The screenshot shows a dialog box titled "Edit Properties" with the following fields:

- Column Name:** DATA00
- Notes:** (Empty text area)
- ID Method:** Auto
- Domain:** DOB
- Algorithm:** sample:MultiColumnDateAl...
- Select Logical Field:** startDate
- Algorithm Group:** 1
- Type:** LOCAL_DATE_TIME
- Date Format:** yyyy-MM-dd

Buttons for "Cancel" and "Save" are located at the bottom right.

i In the UI configuration for columnMetadata, the customer shouldn't mark the *isMasked* field (as via the API in the example above). It's taken care automatically since ME knows the associated logical field is being masked or used as a read-only.

Error Management

There are different configuration errors possible while setting the MC algorithms. The configuration process prevents as many misconfigurations as possible, but some configuration errors can only be detected when a job is executed. For example, if trying to associate a second column to the same (already busy) logical field will result in a configuration error similar to:



In case there is a missed association with the required logical field - that type of error isn't recognized during the configuration, but only during the job execution (which will fail due to that misconfiguration).

Please find below an example of the monitor job error report:

▲ Error Report ✕

Execution Events Learn More		
Event	Cause	Description
JOB_ABORTED	UNHANDLED_EXCEPTION	Exception:Algorithm 'sample:MultiColumnDateAlgorithm' requires 2 fields [endDate, startDate] but found 1 fields [startDate]. Missing fields: [endDate]
75_53.log		
<p>2020-12-14 17:24:16,233 [thread] INFO com.dmsuite.dmsApplicator.masking.XMLGenerator executeMarshalling - Generate request xml started successfully.</p> <p>2020-12-14 17:24:16,782 [thread] INFO com.dmsuite.dmsApplicator.masking.XMLGenerator executeMarshalling - Generate Request xml done successfully.</p> <p>2020-12-14 17:24:16,810 [thread] INFO com.dmsuite.dmsApplicator.masking.transformation.MaskingMarshalling createKettleXML - Generate Transformation XML started successfully</p>		

Limitations for the MC Algorithms

1. Currently, it's possible to run the MC Algorithms only on a single table. Masking multiple tables columns by MC Algorithms is not supported.
2. XML File masking does not support MC algorithms.
3. VSAM File masking does not support MC algorithms.
4. Some types of misconfiguration errors (as described above) are only detected during job execution.

Service interfaces (Algorithms)

Introduction

The Extensible Algorithms framework makes a number of services available to the algorithm implementation. This prevents the algorithm from having to re-implement code to perform certain routine tasks and facilitates seamless integration with the Masking Engine. This functionality is exposed to the algorithm class via the **ComponentService** interface.

Whenever a new Masking Algorithm instance is required for masking, the extensibility framework first injects any saved configuration, then invokes the objects *setup* method. This method is passed a reference to an object that implements **ComponentService**. The algorithm's *setup* method can then use this object to access a number of provider methods:

- *getInstanceName* - Get the name of this instance. Because the instance name it is not typically a configurable field in the algorithm, the *getName* method will not correctly return the name of an algorithm instance, even after JSON configuration injection. This method will always return the correct instance name as known to the Masking Engine.
- *openInputFile* - Access the contents of a file, as described in [this section](#)
- *getAlgorithmByName* - Get a usable instance of another algorithm, as described in [this section](#)
- *getCryptoService* - Access cryptographic methods based on the algorithm's key, as described in [this section](#)
- *getLogService* - Get a logger object, as described in [this section](#)



Getting More Information

Refer to the `com.delphix.masking.api.provider` package in the [Javadoc](#) for detailed information.

Accessing Files

It is often the case that a masking algorithm will require a large library of input values - for example, a set of replacement names or account numbers. In other cases, it may be desirable to store the configuration of a particularly complex algorithm in a format other than JSON, perhaps in order to leverage pre-existing code. To support these use cases, the extensible algorithm framework allows algorithms to access input files from a variety of sources.

Opening Input Files

Algorithms may access files in several locations, both on the engine and over the network. In all cases, access is achieved by invoking the `openInputFile` method of the **ComponentService** object passed to the algorithm's `setup` method to acquire an **InputStream**. The file's location must be specified by an **FileReference** object visible in the object's public fields. This object is passed to the `openInputFile` method. The value of the **FileReference** may be made configurable using the `@JsonProperty` annotation. The `openInputFile` method accepts a URI style syntax that combines standard URL notation for web resources with custom URI types for engine and JAR located files. The following formats are supported for `FileReference` values:

URI Format	Description
<code>http[s]://<>]>]><></code>	To open files located on a remote web server
<code>jar://file/<></code>	To open a file located in the algorithm plugin JAR
<code>delphix-file://upload/</code>	To open a file uploaded using Delphix Masking Engine's <code>fileUpload</code> endpoint. The result of POST ing to the <code>fileUpload</code> API endpoint is a URI in this format that should be used exactly as-is for the <code>uri</code> value of the FileReference .
<code>delphix-file://mount/>>>></code>	To open a file located on a NFS/CIFS mount server that has been mounted inside the Delphix Masking Engine using <code>mountFilesystem</code> endpoint

Example Algorithm

```
public class RedactionFile implements MaskingAlgorithm<String> {
    private String redactionCharacter = null;

    @JsonProperty(value = "file", required = true)
    public FileReference file;
    @Override
    public String getName() {
        return "RedactionFile";
    }

    @Override
    public String mask(@Nullable String input) throws MaskingException {
```

```

    if (input == null) {
        return null;
    }
    StringBuilder returnVal = new StringBuilder();
    for (int i = 0; i < input.length(); i++) {
        returnVal.append(redactionCharacter);
    }
    return returnVal.toString();
}

@Override
public void setup(@NonNull ComponentService serviceProvider) {
    InputStream inputStream = serviceProvider.openInputFile(file);
    try (Scanner scanner = new Scanner(inputStream,
Charset.defaultCharset().name())) {
        redactionCharacter = scanner.nextLine().trim();
    } catch (Exception e) {
        e.printStackTrace();
        throw new RuntimeException("Unable to parse input file", e);
    }
}

@Override
public void validate() throws ComponentConfigurationException {
    GenericReference.checkRequiredReference(file, "file");
}
}

```

Some methods have been omitted for brevity.

This example algorithm is very similar to the `StringRedaction` class discussed earlier, in that it redacts strings by replacing with a same-length string of the redaction character. This variant reads the character to use for redaction from an input file, the location of which is specified in the algorithm's configuration. This is all done in the initialize method:

- The value of the variable `file` is public and marked configurable with `@JsonProperty`.
- The file reference is passed to `serviceProvider.openInputFile` during setup - this allows the algorithm to ingest input files in any supported location.
- The redaction character is read from the input stream and stored in instance variable `redactionCharacter` for use in the `mask` method.
- This class's `validate` method uses the static method `GenericReference.checkRequiredReference` provided by the Masking Plugin API to check the file reference for validity.

Accessing Database Servers (JDBC)

It is often the case that a masking algorithm will require access to a large amount of data such as lookup values for masking input data or storing states of the algorithm. To support these use cases, the extensible algorithm framework allows algorithms to access database servers using JDBC connections.

Opening Database Connection

Algorithms access the database by using [extensible drivers](#). Access is achieved by invoking the `openJdbcConnection` method of the **ComponentService** object passed to the algorithm's `setup` method to acquire a **Connection** (`java.sql.Connection`) object. The file's location must be specified by an **JdbcReference** object visible in the object's public fields. This object is passed to the `openJdbcConnection` method. The value of the **JdbcReference** may be made configurable using the `@JsonProperty` annotation. The **JdbcReference** object requires following fields

1. `jdbcDriverId`: The driver Id of the JDBC Driver uploaded using the JDBC Driver API (`/jdbc-drivers`).
2. `url`: The JDBC URL that will be used to connect to the server. Please don't use this to pass the credentials.
3. `credFileReference`: A **FileReference** object that contains the location of the JSON file that stores the credentials. The schema of the file is:

```
{
  "username": "USERNAME",
  "password": "PASSWORD"
}
```

The file must be a `mount` type **FileReference** object. To see how to create a mount type **FileReference** object, refer to [Accessing Files](#). The **Connection** object is kept open throughout the execution of the algorithm unless it is closed by the algorithm itself.

Example Algorithm

```
public class RedactionDB implements MaskingAlgorithm<String> {
    private String redactionCharacter = null;

    @JsonProperty(value = "jdbc", required = true)
    @JsonPropertyDescription("A reference to a database containing a table
redaction_character")
    public JdbcReference jdbc;

    private static final String GET_REDACTION_CHARACTER = "SELECT redact FROM
redaction_character LIMIT 1";

    @Override
    public String getName() {
        return "RedactionDB";
    }

    @Override
    public String mask(@Nullable String input) throws MaskingException {
        if (input == null) {
```

```

        return null;
    }
    StringBuilder returnVal = new StringBuilder();
    for (int i = 0; i < input.length(); i++) {
        returnVal.append(redactionCharacter);
    }
    return returnVal.toString();
}

@Override
public void setup(@NonNull ComponentService serviceProvider) {
    try (Connection conn = serviceProvider.openJdbcConnection(jdbc);
        PreparedStatement stmt = conn.prepareStatement(GET_REDACTION_CHARACTER)) {
        ResultSet resultSet = stmt.executeQuery();
        List<String> redactionChars = new ArrayList<>();
        if (resultSet.next()) {
            redactionCharacter = resultSet.getString("redact");
        } else {
            throw new RuntimeException("Couldn't find redaction character");
        }
    } catch (SQLException e) {
        throw new RuntimeException(e);
    }
}

@Override
public void validate() throws ComponentConfigurationException {
    GenericReference.checkRequiredReference(jdbc, "jdbc");
}
}

```

Some methods have been omitted for brevity.

This example algorithm is very similar to the `StringRedaction` class discussed earlier, in that it redacts strings by replacing with a same-length string of the redaction character. This variant reads the character to use for redaction from a table in a database, the connection information for which is specified in the algorithm's configuration. This is all done in the initialize method:

- The value of the variable "jdbc" is public and marked configurable with `@JsonProperty`.
- The jdbc reference is passed to `serviceProvider.openJdbcConnection` during setup.
- The redaction character is read from the table `redaction_character` and stored in the instance variable `redactionCharacter` for use in the `mask` method.
- This class's `validate` method uses the static method `GenericReference.checkRequiredReference` provided by the Masking Plugin API to check the jdbc reference for validity.

Algorithm chaining

The extensible algorithm framework allows algorithms to instantiate and call other algorithms. This is useful to allow for the composition and reuse of algorithm behaviors. This feature is referred to as algorithm chaining.

Calling other algorithms

In order to make use of this feature, the caller algorithm must acquire an object of the algorithm class it wishes to call by requesting it by instance name using the *getAlgorithmByName* method of the **ComponentService** object. This is done during the execution of the algorithm's *setup* method.

This method requires that the caller specify two values:

1. A reference to the algorithm instance. This must be stored in an **AlgorithmInstanceReference** object whose value is the name of the algorithm instance. This is *algorithmName* in the Masking API, occasionally referred to as "algorithmCd" or "algorithm code". The **AlgorithmInstanceReference** object must be referenced in a *public* field in the algorithm object.
2. The type of data the returned algorithm object should mask, selected from the core types supported by the extensible algorithm framework. Type adaptation is not currently supported in this context, so the algorithm's native type must be the type requested using *getAlgorithmByName*.

Once an algorithm object has been obtained using *getAlgorithmByName*, a reference to the algorithm object may be kept and that algorithm's *mask* method called as needed.

Examples:

- You are creating algorithm instance A via the Masking API Client Algorithm endpoint, and algorithm A uses *getAlgorithmByName* to find algorithm B during *setup*. For the creation of algorithm A to succeed, algorithm B **must** already exist on the Delphix Masking Engine.
- You are installing a plugin that would create the same algorithm A as a static instance. This will fail if algorithm instance B is not also provided by an algorithm class in the same plugin.

Because it is difficult to predict what algorithm names exist on a Delphix Masking Engine, it is advised that the names of any algorithms used for chaining be supplied in the algorithm's JSON configuration. Hard-coding names of algorithms passed to *getAlgorithmByName* directly in the Java source creates dependencies that are not visible except in the error message that results when the caller algorithm fails to initialize, as described in the second example scenario above. Hard-coded references to other algorithms provided by the same plugin should have the value **":algorithmName"**. The ":" character tells the API to fill in this plugin's name when searching for the instance.

- Algorithm instances provided by plugins (via the *getDefaultInstances* method) are prohibited from having dependencies on algorithm instances provided by other plugins. A way to safely implemented this kind of dependency may be added in the future.

Example algorithm

```
public class RandomizedStringMasking implements MaskingAlgorithm<String> {
    private List<MaskingAlgorithm<String>> algorithmList = new ArrayList<>();
    private Iterator<Integer> randomStream;

    @JsonProperty(value = "algorithmNames", required = true)
    public List<AlgorithmInstanceReference> algorithms;

    @Override
```

```

public String getName() {
    return "Randomized Masking";
}
@Override
public Collection<MaskingComponent> getDefaultInstances() {
    RandomizedStringMasking myInstance =
        new RandomizedStringMasking() {
            @Override
            public String getName() {
                return "Randomized Redaction";
            }
            @Override
            public String getDescription() {
                return "Apply a random redaction algorithm from { X, Y, Z }";
            }
        };
    myInstance.algorithms =
        Arrays.asList(
            new AlgorithmInstanceReference(":Redaction X"),
            new AlgorithmInstanceReference(":Redaction Y"),
            new AlgorithmInstanceReference(":Redaction Z"));

    return Collections.singletonList(myInstance);
}

@Override
public void validate() throws ComponentConfigurationException {
    if (algorithms == null || algorithms.isEmpty()) {
        throw new ComponentConfigurationException(
            "Value for field algorithmNames is missing or empty");
    }
    for (AlgorithmInstanceReference ref : algorithms) {
        GenericReference.checkRequiredReference(ref, "algorithms");
    }
}

@Override
public void setup(@NonNull ComponentService serviceProvider) {
    for (AlgorithmInstanceReference algorithm : algorithms) {
        algorithmList.add(serviceProvider.getAlgorithmByName(algorithm,
MaskingType.STRING));
    }
    randomStream = new Random().ints(0, algorithmList.size()).iterator();
}

@Override
public String mask(@Nullable String s) throws MaskingException {
    return algorithmList.get(randomStream.next()).mask(s);
}

```

Some methods have been omitted for brevity.

This algorithm is configured with a list of other String masking algorithms and masks by calling another algorithm from that list at random. This randomization is not based on the algorithm key, so results will not be consistent across masking runs. In addition, this framework defines a default instance that chooses randomly between algorithms "Redaction X", "Redaction Y" or "Redaction Z" included in the same plugin.

The algorithm's public fields include a list of **AlgorithmInstanceReference** objects, made configurable by the `JsonProperty` annotation.

This algorithm's *setup* method does the following:

- For each algorithm name, it calls *getAlgorithmByName* to instantiate a usable algorithm object, saving them in *algorithmList*.
- It initializes a random number generator to produce integers corresponding to each index in *algorithmList*.

This algorithm's *mask* method selects an algorithm at random from *algorithmList* and calls its *mask* method on the input value, returning the result.

This algorithm's *getDefaultInstances* method creates a single instance that chooses between three algorithms. Each algorithm reference begins with ':', indicating that these algorithms should be found in the same plugin as this algorithm. The *getName* and *getDescription* methods of the returned object are overridden to provide values different from those of the framework itself.

Using cryptographic keys

eCryptography is useful in algorithm development for a range of purposes, from straightforward encryption of value to shuffling collections and permuting data in a manner that is consistent across masking jobs. The extensible algorithm framework automatically provides each algorithm with a cryptographic key. This key is wrapped by a service provider object that implements the **CryptoService** interface, providing a number of useful operations based on the algorithm's key. It is also possible to retrieve the raw key assigned to the algorithm as an array of bytes.

Similar to working with files, there is a **KeyReference** type that represents a reference to the key. This is present to support access to keys stored in alternative locations (ex. a key vault) in the future. Currently, the only supported value for these references is "", which indicates that the per-algorithm key stored on the Delphix Masking Engine should be used.

 When working with the Masking SDK maskApp and maskScript utilities, each algorithm's key is a stable hash of its algorithm name, but maybe temporarily set to a random value using the -K flag.

Using the CryptoService provider

The first step any algorithm that wishes to use its algorithm key must take is to retrieve a handle to a cryptographic service provider during initialization. This is done by calling the **ComponentService** object's *getCryptoService* method. The returned provider wraps the key. The operations supported by the **CryptoService** interface are as follows:

- *getRawKey* - retrieve the raw key associated with this provider as a byte array.
- *wrap* - wraps an array of bytes to create a CryptoService object. This is useful for accessing CryptoService methods when the algorithm's key is stored in an alternative location or hard-coded in the algorithm source.
- *deriveNewKey* - derive a new key by permuting this provider's key using SHA-256. A new CryptoService object is returned wrapping the new key. The zero-argument version of this method returns the same key each time it is called on the same provider - in order to create multiple, different keys, a different salt must be provided to each method call. It is advisable that whenever an algorithm wishes to use cryptography for multiple purposes, new and distinct keys be derived for each purpose.
- *computeHashedLookupIndex* - compute an integer value from 0 to (modulus - 1) by hashing the input value + key. This method is designed to allow randomized, but consistent, lookups into a replacement table based on the input value.
- *shuffleList* and *shuffleListNoCollisions* - these methods shuffle their argument **List** in-place using the key to seed the randomization. The "noCollisions" variant ensures that no object in the list remains in its original position.

Example algorithm

```
public class StringHashedLookup implements MaskingAlgorithm<String> {
    private List<String> replacements;
    private CryptoService crypto;

    public KeyReference key = new KeyReference();

    @JsonProperty("replacementFile")
    public FileReference replacementFile;

    @Override
```

```

public void validate() throws ComponentConfigurationException {
    GenericReference.checkRequiredReference(replacementFile, "replacementFile");
}

@Override
public void setup(@NonNull ComponentService serviceProvider) {
    replacements = new ArrayList<>();

    String line;
    try (InputStream is = serviceProvider.openInputFile(replacementFile);
        BufferedReader reader =
            new BufferedReader(new InputStreamReader(is, "UTF_8"))) {
        while ((line = reader.readLine()) != null) {
            replacements.add(line);
        }
    } catch (IOException e) {
        throw new RuntimeException(e);
    }
    crypto = serviceProvider.getCryptoService(key);
}

@Override
public String mask(@Nullable String input) {
    if (input == null || input.length() == 0) {
        return input;
    }
    return replacements.get((int) crypto.computeHashedLookupIndex(input,
replacements.size()));
}
}

```

Some methods have been omitted for brevity.

This example algorithm functions very similarly to the existing Secure Lookup algorithm, except it employs a different hash method from the new **CryptoService** provider.

- The algorithm is configured with an input file by supplying a public, annotated **FileReference** field *replacementFile*.
- In the *setup* method, the replacement file is ingested and saved as a list of values.
- Additionally in *setup*, the cryptographic service provider is initialized using the default key reference, accessing the algorithm's key.
- The mask method uses the *computeHashedLookupIndex* method to compute the index of the replacement to use from the *replacements* list.

Logging

It is possible for a plugin algorithm to write information into the job logs, and consequently, Continuous Compliance Engine logs. This is accomplished by using calling the `getLogService` method of the **ServiceProvider** interface provided at the algorithm setup. The resulting **LogService** object may be used to make logging entries at various levels of severity. The available log levels are ERROR, WARNING, INFO, and DEBUG.

The log interface is provided to allow for debugging output during algorithm development, and for reporting of statistical or similar values detailing the overall operation of the algorithm, typically in the `tearDown` method.

i Logging Security Warning
An algorithm **must** never log unmasked values (the `input` argument to the `mask` method) to the log files. The job and Masking Engine log files may be retrieved by engine users and are included support bundles.

i Logging Verbosity
Algorithms also should not log progress messages or other verbose details, especially from the `mask` method, as this will fill the log files with messages and may impact job performance. There is a rate-limiting mechanism that limits the volume of messages each algorithm can write over time, but any amount of routine logging is likely to diminish the overall usefulness of the logs by obscuring more important messages.

Example code

This example is take from the StringRedaction sample algorithm provided with the SDK:

```
public class StringRedaction implements MaskingAlgorithm<String> {
    ...
    private LogService logger;
    ....

    @Override
    public String mask(@Nullable String input) throws MaskingException {
        if (input == null) {
            return null;
        }

        if (random.nextDouble() < 0.1) {
            logger.info("{0}: Masked {1} values", getName(), count);
        }

        StringBuilder returnVal = new StringBuilder();

        for (int i = 0; i < input.length(); i++) {
            returnVal.append(redactionCharacter);
        }
        count++;
        return returnVal.toString();
    }

    @Override
    public void validate() throws ComponentConfigurationException {
```

```
        if (redactionCharacter == null || redactionCharacter.length() != 1) {
            throw new ComponentConfigurationException(
                "redactionCharacter must be a single character");
        }
    }

    @Override
    public void setup(@NonNull ComponentService serviceProvider) {
        logger = serviceProvider.getLogService();
    }

    @Override
    public void tearDown() {
        logger.info("{0}: Masked a total of {1} values", getName(), count);
        count = 0;
    }
}
```

Some methods and fields elided for the sake of brevity

The relevant details here:

- The *setup* method uses the provided **ComponentService** object to get a **LogService** instance, saving it as *logger*.
- The *mask* method calls the logger's *info* method to write informational messages at random during execution. This kind of "progress" logging may be useful during development but should be removed for algorithms before production deployment.
- The *tearDown* method calls the *info* method of the logger again to record the total number of values masked.

Security considerations

It is important that only well-crafted and trustworthy plugin modules are installed on the Continuous Compliance Engine; otherwise, the security of the appliance and masked data may be compromised. This section contains information for developers on how to ensure that their algorithm plugins function securely, as well as for engine administrators to ensure that only trusted plugins are installed and executed on the engine.

Algorithm implementation

This section details a number of security considerations developers should be aware of when creating plugin algorithms for the Continuous Compliance Engine.

The security sandbox

During execution, all plugin code is sandboxed using the Java Security Manager. Plugins are granted all permissions *except* for the following non- `FilePermission` :

Class	Target	Action
<code>java.net.SocketPermission</code>	<code>localhost:-</code>	accept, connect, listen, resolve
<code>java.lang.RuntimePermission</code>	<code>exitVM</code>	
<code>java.lang.RuntimePermission</code>	<code>createClassLoader</code>	
<code>java.lang.RuntimePermission</code>	<code>accessClassInPackage.sun</code>	
<code>java.lang.RuntimePermission</code>	<code>setSecurityManager</code>	
<code>java.security.SecurityPermission</code>	<code>setPolicy</code>	
<code>java.security.SecurityPermission</code>	<code>setProperty.package.access</code>	

With regards to `FilePermissions`, `read` access is granted to all, though `write` is only allowed for the following directories:

- the masking user's home directory (`System.getProperty("user.home")`)
- the JVM's default temp directory (`System.getProperty("java.io.tmpdir")`)

Please note that both of these locations are shared, so care will need to be taken to avoid collisions.

The set of permissions granted to plugins is static and cannot be modified. To facilitate testing, the same security restrictions are applied when plugins are run using the `maskApp` or `maskScript` utilities in the Masking SDK (with the exception of the `SocketPermission` and all instances of `write FilePermission`).

Handling errors

One important aspect of ensuring that an algorithm securely masks sensitive data is proper handling any errors that might occur during algorithm execution.

One particular category of error that might occur is when the input value does not match the format expected by the algorithm. Perhaps an account number masking algorithm is applied to a column containing free-text comments, or an image blurring algorithm is applied to non-image binary data. This is referred to as Non-conformant data. The Algorithm Extension Plugin API defines how an algorithm may trigger the Non-conformant data handling mechanisms built into the Masking Engine.

Reporting non-conformant data

Whenever a Non-conformant input value is encountered, and the algorithm cannot mask it, the algorithm *mask* method should throw an exception of class **NonConformantDataException** supplied by the Masking Plugin API. This triggers the Non-conformant data reporting mechanism of the masking engine. The **String** value used to construct this exception **must not** include the unmasked input value, as this would result in the sensitive value being saved in the Masking Engine logs and made visible in the engine UI. A redacted sample of the Non-conformant data will be saved automatically by the reporting mechanism.

Handling other errors

In general, other code errors should be handled as responsibly as possible by the algorithm implementations, following these guidelines:

- Under no circumstances should the unmasked input values (the *input* argument to the *mask* method) be included in any **Exception** thrown. Exception details are recorded in the engine logs, making them visible to the engine operator and subject to potential disclosure in support bundles. Similarly, exceptions should not simply be re-thrown as **NonConformantDataException** as the original exception's message may contain the sensitive value.
- Whenever possible, configuration problems should be reported in the *validate* or *setup* method, rather than the *mask* method. Waiting until the *mask* method has run to report an error allows the masking job to run, potentially leaving the database table or file partially masked.
- An algorithm should **never** fail in such a way that sensitive values pass through without being masked. In such cases, non-conformant can be reported as described above.

Logging

The extensibility framework provides the capability for an algorithm to create a [logger](#) in order to write diagnostic messages to the Continuous Compliance Engine logs. **Under no circumstances should unmasked data (any input argument values to the *mask* method) be logged.** Logged messages are visible to users via the UI and web API, and may be disclosed in support bundles. It is recommended that production algorithms never log in the *mask* method, for both performance and security reasons.

Additionally, plugin code should *never* read or write any of the *System* input or output streams. Specifically, these are *System.in*, *System.out*, and *System.err*. All logging should be done using the provided logging interfaces.

Handling secret credentials and keys

The JSON document describing the configuration of each algorithm is stored unencrypted on the Continuous Compliance Engine and made visible to users with access privileges through the UI and web API. For these reasons, secret values of any kind should **never** be part of an algorithm's configuration, regardless of whether the algorithm is user-created or built into a plugin. This includes secret keys, as well as access credentials or API keys that might be used to access remote systems. The only mechanism available as of release 6.0.3.0 that would allow an algorithm to load a sensitive value without the risk of compromise is reading the value from a file stored on an NFS or CIFS [mounted filesystem](#).



A feature to allow plugins to securely access managed credentials will be added in a future Delphix release.

Secret values (keys) or seeds that drive the output "randomization" an algorithm should not be embedded in the algorithm code. Instead, the algorithm's assigned key should be accessed via the [CryptoService interface](#). Static secrets of this kind of risk disclosure should the plugin JAR file be disclosed. There are also risks associated with the algorithm producing the same masking results in all cases, especially if the plugin is to be used for masking by multiple organizations.

Driver supports

Introduction

As of release 6.0.9.0, the Continuous Compliance Engine supports the installation of driver support plugins, written in Java, that provide tasks to execute before/after masking jobs on extended database connectors. Note that this feature requires creating/updating an uploaded JDBC driver to reference the driver support plugin, which is only possible via the web API. Thus creating an extended database connector using that JDBC driver and a corresponding masking job will allow you to enable whatever available tasks that are implemented by the driver support on the job, which you can do via the web API and UI. This process is detailed further [here](#). This feature is referred to as Extensible Driver Supports. This section of the documentation details all aspects of masking driver support plugin usage and development. The *Guided Tour* portion of the [workflows section](#) walks the user through the basic process of building a simple plugin and installing it onto the Continuous Compliance Engine. Other sections explore topics such as the [DriverSupport interface](#) and [service interface](#).

This documentation assumes the reader has some familiarity with Java development as well as operation of the Delphix Masking Engine via both the UI and Web API Client. The reader should also understand the security requirements associated with any new driver supports being developed.

SDK features

The Extensible SDK provides a number of useful functions that aid development of new driver supports for the Continuous Compliance Engine. It is available on the Delphix software [download site](#).

- Creation of empty "skeleton" projects, with build files - the maskScript *init* sub-command
- Testing of the execution of driver support tasks on a database without a masking engine
 - The maskScript *taskExecute* sub-command (**NOTE:** If you want to verify that the **preJobExecute** part of the task was successfully executed, you will want to comment out the reversal of the task in **postJobExecute**, or vice versa. Otherwise, set up your [development environment](#), add a breakpoint and use the debugger to pause after **preJobExecute** execution.)
- Uploading of plugins to the masking engine - the maskScript *install* sub-command
- Sample driver support for MSSQL extended database connector

Getting more information

Several other sources of information are available to aid in plugin development:

- The <http://README.md> file under docs in the Extensible SDK download archive
- The [Masking Plugin API Javadoc](#)
- Invoke **maskScript** (located under *sdkTools/bin* in the SDK download) with the -h option for usage help

The DriverSupport Java interface

Any Java class that should be recognized as a driver support plugin must implement the **DriverSupport** interface. The full details of this interface are described in the [Masking Plugin API Javadoc](#).

Method overview

This section provides a high-level overview of the methods in the **DriverSupport** interface. For complete details, consult the Masking Plugin API Javadoc included in the Algorithm SDK archive.

- *getTasks* - This method is used to determine the list of available tasks to execute on a corresponding data source. The order in which the tasks are added to the list of tasks indicates the order in which the tasks will be executed on the target data source.

The life cycles of driver support objects

The Extensibility framework uses objects classes implementing **DriverSupport** interface for several distinct purposes. These object life cycles are as follows:

Plugin discovery

This occurs when the extensibility framework evaluates the capabilities present in a **DriverSupport** class.

1. Java object creation - an object of the driver support class is created
2. *getTasks*- determines all available tasks
 - *getTaskName* - get the name of each task
3. Disposal - the Java object is discarded

Driver support use

This is the life cycle of a driver object when executing a masking job.

1. Java object creation - an object of the driver support class is created
2. Configuration injection - the masking inventory is used to instantiate a JobInfo object and the database connection is used to instantiate the Connection object (the target SQL connection)
3. *setup* - the *setup* method is called once
4. *preJobExecute* - the *preJobExecute* method is called once before executing the transformation
5. *postJobExecute* - the *postJobExecute* method is called once after executing the transformation
6. Disposal - the Java object is discarded

SDK workflows (Driver supports)

Introduction

This section is intended to walk a developer through several workflows using the Delphix Extensible SDK, such as creating a new algorithm or driver support plugin and installing it on a Continuous Compliance Engine.

In order to develop and deploy driver support plugins, you will interact primarily with two tools - the Masking API client, and the Masking Extensible SDK. The Masking API client is a long-standing feature that allows interactive execution of API operations on the Continuous Compliance Engine, while the Masking Extensible SDK is a software package created specifically to aid in driver support development.

Outline for a guided tour

By following the steps in the outline below, you can tour the basic functionality provided by the Extensible Driver Support feature and Extensible SDK.

1. Create a driver support plugin by choosing one of two options:
 - a. [Building the sample driver support project](#)
 - b. [Creating and building your own driver support project](#)
2. [Test the driver support plugin using maskScript](#)
3. [Install the newly created plugin on the Continuous Compliance Engine](#)
4. [View and manage the plugins on a Continuous Compliance Engine using the API Client](#)

Building the sample plugin (SDK workflows/Driver supports)

The Extensible SDK contains a buildable Sample Driver Support Plugin with a functional driver support illustrating the features of the Extensibility Framework. These simple commands build the plugin containing the sample driver support.

Starting from ***sdk_root***:

```
$ cd samples
$ ./gradlew :driverSupport:jar
```

This creates the Sample Driver Support plugin JAR file ***sdk_root***/samples/build/libs/driverSupport.jar.

The Sample Driver Support project provides a convenient way to see a working example plugin.

i While it is possible to modify these driver supports by changing the Java source and rebuilding the plugin, when starting a new project to develop one, it is highly recommended that you [create your own project](#) rather than modifying files in the Sample Driver Support project subtree. This will prevent the loss of customizations to the project build files should you chose to install a new version of Masking Extensible SDK over your existing SDK directory.

Creating a new project (SDK workflows/Driver supports)

This section describes how to create a brand new Java project for a new masking driver support plugin. We will use the `maskScript` utility to create a skeleton project and an empty driver support class in that project.

Creating the project

Before you begin, you'll want to pick a name for your project, and an **empty** directory (outside of the Masking SDK source tree) where your project will be created. Once you've done this, run this **maskScript** command:

```
$ maskScript init -t driverSupport -d <project path> -n <project name> -a <author name> -v <version>
```

For example, this command will create a project named `demoProject` in the `demo-proj` subdirectory of your home directory.

```
$ maskScript init -t driverSupport -d $HOME/demo-proj -n demoProject -a <plugin author's name> -v <version>
```

For the rest of this section, we'll assume a new project has been created under **proj_dir**. Change your working directory to **proj_dir**. You'll notice that the project is created with a sample driver support file `proj_dir/src/main/java/com/sample/masking/driverSupport/MSSQLDriverSupport.java`. It's possible to build this into a usable plugin by running:

```
$ cd <proj_dir>
$ ./gradlew jar
```

 This sample driver support project is not intended to be used in a production environment and is only meant to serve as an example

Creating a driver support class

Run the `maskScript` utility to create a skeleton class file:

```
$ cd <proj_dir>
$ maskScript generate driverSupport -p com.delphix.demo -c <class_name> -s .
```

By convention, the class file `.java` will be created under a sub-directory path based on the package name, so it might be helpful to use the `find` command to locate it:

```
$ find . -name <class_name>.java
./src/main/java/com/delphix/demo/<class_name>.java
```

The initial content of this file is:

```
package com.delphix.demo;
```

```

import com.delphix.masking.api.driverSupport.DriverSupport;
import com.delphix.masking.api.driverSupport.Task;
import com.delphix.masking.api.driverSupport.jobInfo.JobInfo;
import com.delphix.masking.api.provider.ComponentService;
import com.delphix.masking.api.provider.LogService;
import java.sql.Connection;
import java.util.ArrayList;
import java.util.List;

public class <class_name> implements DriverSupport {

    /**
     * This method serves as a directory of Task objects provided by this plugin.
     *
     * @return an ordered list of tasks. The order that tasks are added to the
returning list is the
     *     order that they will be executed in.
     */
    @Override
    public List<Task> getTasks() {
        // TODO: return list of implemented task objects
        List tasks = new ArrayList<>();
        tasks.add(new ExampleTask());

        return tasks;
    }

    public class ExampleTask implements Task {
        private JobInfo jobInfo;
        private LogService logService;
        private Connection targetConnection;

        @Override
        public String getTaskName() {
            return "Example Task";
        }

        @Override
        public void setup(ComponentService serviceProvider) {
            this.jobInfo = serviceProvider.getJobInfo();
            this.targetConnection =
serviceProvider.getTargetConnection();
            this.logService = serviceProvider.getLogService();
        }

        @Override
        public void preJobExecute() {
            // TODO: implement code to execute BEFORE masking job runs.
        }

        @Override

```

```

        public void postJobExecute() {
            // TODO: implement code to execute AFTER masking job runs.
        }
    }
}

```

Implementing the driver support class

The first thing to notice about the skeleton driver support class is that the `getTasks` method just returns an array of tasks with a single no-op task called `ExampleTask`. This means no actual additional transaction will be performed on the target data as part of a masking job, so this will certainly need to change.

It is recommended that you change the task class to a name that more accurately reflects what the task does as well as the string returned from the method `getTaskName`. Delete the `TODO` comments in

In order to rebuild the project to generate the driver support plugin JAR, you'll need to first update `settings.gradle` to include the project directory:

```

/*
 * Copyright (c) 2019, 2021 by Delphix. All rights reserved.
 */

pluginManagement {
    resolutionStrategy {
        eachPlugin {
            if ( requested.id.id == 'com.diffplug.gradle.spotless' ) {
                useModule( "com.diffplug.spotless:spotless-plugin-
gradle:$spotlessVer" )
            }
        }
    }
}

rootProject.name = '<proj_dir>'
include 'sdkTools'
include 'algorithm'
include 'assemble'
include 'driverSupport'

```

Then to generate the driver support plugin JAR:

```
$ ./gradlew jar
```

This creates or updates the plugin JAR file `proj_dir/build/libs/.jar`

Service discovery (SDK workflows/Driver supports)

Java service discovery is used to determine which classes in the plugin JAR present relevant functionality to the Delphix Masking Engine. When a plugin is loaded, the file *com.delphix.masking.api.plugin.DriverSupport* under *META-INF/services* in the JAR is consulted for a list of classes that implement the **DriverSupport** interface.

- When the maskScript *generate* sub-command is used to create a new driver support class, the service discovery metadata file is automatically updated.

If a driver support class is missing from the services file, it will not be usable when the plugin is loaded. It is essentially invisible to the extensibility framework. If a class is mentioned in this file but not present in the JAR, the plugin will fail to load.

Executing a driver support task using the SDK (SDK workflows/Driver supports)

It will often be more convenient to use the SDK utilities to test a driver support since this avoids the need to install or update your plugin, create or update a jdbc driver to reference the driver support plugin, and execute jobs on the Delphix Masking Engine. This can be done from the command line using maskScript.

Using maskScript to test a driver support task

The **maskScript** utility is non-interactive, which lets you execute a task on a given data source. The jdbc driver, driver support and task are selected using command-line options. This example uses the Sample Driver Support plugin. This plugin can be built using the process described [here](#).

Create a task set up json file that corresponds to the specific table and desired database with the contents:

```
{
  "tableMetadata": [
    {
      "name": "Person",
      "schema": "dbo",
      "columns": [
        {
          "name": "column_pk"
        },
        {
          "name": "column_name_1"
        },
        {
          "name": "column_name_2"
        },
        {
          "name": "column_name_3"
        }
      ]
    }
  ],
  "jdbcConnection": {
    "username": "USERNAME",
    "password": "PASSWORD",
    "host": "jdbc:sqlserver://HOST:1433;databaseName=DB_NAME",
    "propertyFilePath": ""
  }
}
```

Execute the task by indicating the name of the desired task, driver support filepath, task set up json, and jdbc driver:

```
$ maskScript taskExecute -n "Task Name" -j /path/to/driverSupport.jar -c /path/to/task-setup.json -l /path/to/jdbcDriver.jar
```

 In order to be usable, the class that implements **DriverSupport** must also be listed in the appropriate service description file. Refer to [this section](#) for details.

Use any available database management tool like [DbVisualizer](#) to connect to the database and verify that the task was successfully executed.

Retrieving information about installed plugins (SDK workflows/Driver supports)

The GET endpoints are useful for getting information about plugins. After following the steps in [this section](#) to install the Sample Driver Support plugin, the GET operation will return (elided for brevity):

```
{
  "pluginId": 9,
  "pluginName": "Sample Plugin",
  "pluginAuthor": "Sample Plugin Author",
  "pluginType": "DRIVER_SUPPORT",
  "originalFileName": "driverSupport.jar",
  "originalFileChecksum":
"f8398c0768ecf7709c6992b3f048f9da8be640285b3ccc968973949ca3cceb02",
  "installDate": "2021-04-21T15:29:01.982+00:00",
  "installUser": 5,
  "builtIn": false,
  "pluginVersion": "1.5.0",
  "pluginObjects": [
    {
      "objectIdentifier": "1",
      "objectName": "Disable Constraints",
      "objectType": "DRIVER_SUPPORT_TASK"
    },
    {
      "objectIdentifier": "2",
      "objectName": "Disable Triggers",
      "objectType": "DRIVER_SUPPORT_TASK"
    },
    {
      "objectIdentifier": "3",
      "objectName": "Drop Indexes",
      "objectType": "DRIVER_SUPPORT_TASK"
    }
  ]
},
...
```

i The `objectIdentifier` field refers to the ID of the task. The order in which the tasks are returned from the API is the order in which the tasks will be executed; the `objectIdentifier` (task ID) has no bearing on the task execution order.

For each plugin, the plugin metadata, including `pluginId`, `pluginName` and `originalFileChecksum` are displayed first. This is followed by a list of tasks included in the plugin.

Service Interface (Driver supports)

Introduction

The Extensible Driver Supports framework makes certain services available to the driver support implementation. This prevents the driver support from having to re-implement code to perform certain routine tasks and facilitates seamless integration with the Masking Engine. This functionality is exposed to the driver support class via the **ComponentService** interface.

Whenever a new Masking driver support instance is required for masking, the extensibility framework first injects any saved configuration, then invokes the objects *setup* method. This method is passed a reference to an object that implements **ComponentService**. The driver support's *setup* method can then use this object to access a number of provider methods:

- *getInstanceName* - Get the name of this instance. Because the instance name it is not typically a configurable field in the driver support, the *getName* method will not correctly return the name of an driver support instance, even after JSON configuration injection. This method will always return the correct instance name as known to the Masking Engine.
- *getTargetConnection* - Gets a `java.sql.Connection` that is made using the target database connector.
- *getJobInfo* - Gets a `jobInfo` object, which maps the names of tables, schemas, and columns that are in the masking ruleset.
- *getLogService* - Get a logger object, as described in [this section](#)



Getting more information

Refer to the `com.delphix.masking.api.provider` package in the [Javadoc](#) for detailed information.

Accessing masking engine rulesets

The `JobInfo` object represents the database connector's inventory on the masking engine. It contains all of the columns that are going to be masked along with the table and/or schema that they belong to.

Example driver support task

```
public class DropIndexes implements Task {
    private JobInfo jobInfo;
    private LogService logService;
    private Connection targetConnection;

    @Override
    public String getTaskName() {
        return "Drop Indexes";
    }

    @Override
    public void setup(ComponentService serviceProvider) {
        this.jobInfo = serviceProvider.getJobInfo();
        this.targetConnection = serviceProvider.getTargetConnection();
        this.logService = serviceProvider.getLogService();
    }

    /**
     * This method is to structure all of the columns belonging to the jobInfo.
     *
     * @return A String of comma separated column names.
     */
    private String getCommaSeparatedColumnNames() {
        StringBuilder resultStringBuilder = new StringBuilder();
        for (TableInfo table : jobInfo.getTables()) {
            resultStringBuilder.append(
                table.getColumns().stream()
                    .map(ColumnInfo::getName)
                    .map(this::singleQuoted)

                .collect(Collectors.joining(",")));
            resultStringBuilder.append(",");
        }
        String commaSeparatedResult = resultStringBuilder.toString();
        return commaSeparatedResult.substring(0,
            commaSeparatedResult.length() - 1);
    }
}
```

Some methods have been omitted for brevity.



See the [Javadocs](#) for further information on the `JobInfo`, `SchemalInfo`, `TableInfo` and `ColumnInfo` interfaces.

Accessing database server (JDBC)

Driver support plugins will require access to the target database table on which its selected tasks will be run as part of a masking job. The extensible driver support framework allows driver supports to access database servers using JDBC connections, utilizing the existing masking web API. The same connection that is built during the test connection endpoint (`POST /database-connectors/{connector_id}/test`) on the masking engine is the same connection that will be returned by the service provider's **getTargetConnection** method.

Example driver support task

```
public class DisableTriggers implements Task {
    ...
    private Connection targetConnection;
    ...

    @Override
    public String getTaskName() {
        return "Disable Triggers";
    }

    @Override
    public void setup(ComponentService serviceProvider) {
        this.jobInfo = serviceProvider.getJobInfo();
        this.targetConnection = serviceProvider.getTargetConnection();
        this.logService = serviceProvider.getLogService();
    }

    ...

    @Override
    public void preJobExecute() throws MaskingException {
        long start = System.currentTimeMillis();
        this.triggersOnMaskedTables = findEnabledTriggersOnMaskedTables();
        try (Statement statement = targetConnection.createStatement()) {
            for (Map.Entry<String, String> entry : triggersOnMaskedTables.entrySet())
            {
                String triggerName = entry.getKey();
                String tableName = entry.getValue();
                String disableTriggersStatement =
                    String.format(MODIFY_TRIGGERS_SQL, "DISABLE", triggerName,
tableName);
                try {
                    statement.execute(disableTriggersStatement);
                } catch (SQLException e) {
                    String errorMessage = ...;
                    logService.error(errorMessage + e);
                    throw new MaskingException(errorMessage, e);
                }
            }
        } catch (SQLException e) {
```

```
String errorMessage = "Error creating a statement on target connection.";
logService.error(errorMessage + e);
    throw new MaskingException(errorMessage, e);
}
}
}
```

Some methods have been omitted for brevity.

Logging (Service interfaces)

It is possible for a driver support plugin to write information into the app logs, and consequently, Continuous Compliance Engine logs. This is accomplished by using calling the `getLogService` method of the **ServiceProvider** interface provided at the driver support setup. The resulting **LogService** object may be used to make logging entries at various levels of severity. The available log levels are ERROR, WARNING, INFO, and DEBUG.

The log interface is provided to allow for debugging output during driver support development, and for reporting of statistical or similar values detailing the overall operation of the driver support, typically in the `tearDown` method.

i **Logging Verbosity**
Driver support tasks also should not log progress messages or other verbose details, especially from the **preJobExecute** or **postJobExecute** methods, as this will fill the log files with messages and may impact job performance. There is a rate-limiting mechanism that limits the volume of messages each driver support can write over time, but any amount of routine logging is likely to diminish the overall usefulness of the logs by obscuring more important messages.

Example code

This example is take from the MSSQL sample Disable Constraints driver support task provided with the SDK:

```
public class DisableConstraints implements Task {
    ...
    private LogService logService;
    ...

    @Override
    public String getTaskName() {
        return "Disable Constraints";
    }

    @Override
    public void setup(ComponentService serviceProvider) {
        this.jobInfo = serviceProvider.getJobInfo();
        this.targetConnection = serviceProvider.getTargetConnection();
        this.logService = serviceProvider.getLogService();
    }

    ...

    @Override
    public void preJobExecute() throws MaskingException {
        long start = System.currentTimeMillis();
        disableConstraints();
        logService.info(
            String.format(
                "Total execution to disable all constraints on masked tables
took %s ms.",
                String.valueOf(System.currentTimeMillis() - start)));
    }
}
```

```

/** This function enables all constraints on the target database table. */
private void disableConstraints() throws MaskingException {
    this.enabledConstraints = findEnabledConstraints();
    try (Statement statement = targetConnection.createStatement()) {
        for (ConstraintMetadata constraint : enabledConstraints.values()) {
            logService.info(
                String.format(
                    "Starting to disable constraint: \"%s\" on table
                    \"%s\"",
                    constraint.getName(),
                    constraint.getQualifiedTableName()));
            try {
                String builtSqlStatement =
                    String.format(
                        ALTER_CONSTRAINT_STATEMENT,
                        constraint.getQualifiedTableName(),
                        constraint.getDisableAction(),
                        constraint.getName(),
                        "");
                logService.info(builtSqlStatement);
                statement.execute(builtSqlStatement);
            } catch (SQLException e) {
                String errorMessage =
                    String.format(
                        "Error disabling constraint: \"%s\" on table
                        \"%s\".",
                        constraint.getName(),
                        constraint.getQualifiedTableName());
                logService.error(errorMessage + e);
                throw new MaskingException(errorMessage, e);
            }
            logService.info(
                String.format(
                    "Finished disabling constraint: \"%s\" on table
                    \"%s\".",
                    constraint.getName(),
                    constraint.getQualifiedTableName()));
        }
    } catch (SQLException e) {
        String errorMessage =
            String.format(
                "Error creating statement on target connection %s: ",
                targetConnection.getClass());
        logService.error(errorMessage + e);
        throw new MaskingException(errorMessage, e);
    }
}
...

@Override
public void postJobExecute() throws MaskingException {
    long start = System.currentTimeMillis();

```

```
enableConstraints(); // comment this out if testing of the task execution via
the SDK is desired
logService.info(
    String.format(
        "Total execution to enable all constraints on masked tables
took %s ms.",
        System.currentTimeMillis() - start));
}
```

Many methods and fields elided for the sake of brevity

The relevant details here:

- The *setup* method uses the provided **ComponentService** object to get a **LogService** instance, saving it as *logService*.
- The *disableConstraints* method calls the logger's *info* method to write informational messages at random during execution. This kind of "progress" logging may be useful during development but should be removed for driver supports before production deployment. It also calls the logger's *error* method in the event of a failure to connect to the data source or otherwise execute the task on the given data source.

Managing plugins using the API client

The Continuous Compliance Engine's web API includes a *plugin* endpoint for managing plugins:

plugin		Show/Hide	List Operations	Expand Operations
GET	/plugin			Get all plugins
POST	/plugin			Install plugin
DELETE	/plugin/{pluginId}			Delete plugin
GET	/plugin/{pluginId}			Get plugin detail by pluginId
PUT	/plugin/{pluginId}			Update plugin

Displaying information about installed plugins

The GET endpoints are useful for getting information about plugins. After following the steps in [this section](#) to install the plugin, the GET operation will allow you to retrieve information about the installed plugins. To know what response and information to expect, please see the respective documentation for [driver supports](#) and [algorithms](#).

Other plugin endpoint operations

In addition, to GET, the *plugin* endpoint supports the other CRUD operations:

- POST - install a new plugin
- PUT - update an existing plugin
- DELETE - remove a plugin from the system

The POST and PUT operations both require a *fileReference* value representing the plugin file to be installed or updated. These values are the result of using the *fileUpload* endpoint to upload the plugin JAR file to the Masking Engine.

In order to install a new version of this plugin, one could use the PUT operation, or, assuming the algorithm or driver support plugin are not in use, simply DELETE the plugin and POST a new version (or install using the SDK maskScript). Both PUT and DELETE operations require the pluginId value listed for each plugin using the GET operation. Refer to [this section](#) for details to help the plugin author ensure that new versions of a plugin can successfully install over an existing version using the PUT operation.

Installing a plugin onto the Delphix masking engine

Once you've successfully built a plugin, it's possible to upload it using the *fileUpload* endpoint in the Masking Engine's API Client, then install the plugin using the *plugin* endpoint. The SDK's **maskScript** includes a sub-command to automate this process. Replace "admin" with your username if you prefer to install the plugin as another user.

```
$ maskScript install -j <path to plugin JAR> -H <engine hostname> -u admin
```

For example, if you've chosen to build the included Sample Algorithm Plugin in its standard location, and the IP address of your Delphix Masking Engine is 10.0.0.1, this command would install the Sample Algorithm Plugin onto your engine:

```
$ maskScript install -j algorithm/build/libs/algorithm.jar -H 10.0.0.1 -u admin
```

You will be prompted for the Delphix Masking Engine user's password.

Upon success, this command will display the JSON response from the API request, including details about the installed plugin as well as a list of the frameworks and algorithms that were installed.

When installing a plugin using the **maskScript**, the *-n* option may be used to override the plugin name on the Masking Engine. This may be used to install two plugins with the same built-in name on the engine at once (for example, two different versions of the same plugin), but should usually be avoided due to the potential confusion that can result from installing the same plugin on multiple engines with different names.



The Web API Client may also be used to manage the plugins installed on the Delphix Masking Engine, as described in this [section](#). Also, algorithms support installing [multiple plugins](#) on a masking engine.

Secure plugin deployment

It is absolutely vital that only known plugin modules from trusted vendors be installed on the Delphix Masking Engine. A bad plugin may include algorithms that malfunction, possibly by failing to mask data or entering a loop consuming CPU or memory resource. This can lead to job failure, the engine UI becoming unresponsive, or failure to properly mask sensitive data in the case of [algorithms](#). Plugin execution is sandboxed using the Java Security Manager to guard against malfunctioning code. However, JVM security has historically proven susceptible to allowing untrusted modules to run with the danger of malicious code gaining enhanced or full access to the system running the JVM.

With these considerations in mind, this section describes steps the Delphix Masking Engine administrator can take to ensure that only trusted plugins are executed.

Using roles to restrict plugin installation

This [section](#) describes how to define roles and assign roles to Delphix Masking Engine users. The new profile privilege **Plugins** controls which users are able to install new plugins on to the engine. It is advised that only users that **need** the ability to install plugin modules onto the engine be granted roles that include this privilege.

Verifying the SHA256 hash of installed plugins

When the Masking Web API Client *plugin* endpoint is used to GET the details of a plugin, the field *originalFileChecksum* contains the SHA256 hash of the plugin file installed. This may be compared to a vendor-supplied list of known plugin hashes to verify that a plugin installed on the Delphix Masking Engine has not been tampered with.

For example:

```
{
  "pluginId": 9,
  "pluginName": "demoPlugin",
  "originalFileName": "demoProject.jar",
  "originalFileChecksum":
  "65053d20874ec7929d219b24bdf98ac5b6f7b06ac6bab59712cf78971be135c9",
  "installDate": "2020-06-24T18:19:42.534+0000",
  "installUser": 5,
  "builtIn": false,
  "pluginVersion": "1.0.0",
  "pluginObjects": [
    {
      "objectIdentifier": "demoPlugin:Clobber",
      "objectName": "demoPlugin:Clobber",
      "objectType": "ALGORITHM"
    },
    {
      "objectIdentifier": "demoPlugin:SampleAlgorithm",
      "objectName": "demoPlugin:SampleAlgorithm",
      "objectType": "ALGORITHM"
    }
  ]
}
```

Most UNIX like operating systems provide a way to compute the same hash of a file on the command line.

Apple OSX Example:

```
$ shasum -a 256 demoProject.jar  
65053d20874ec7929d219b24bdf98ac5b6f7b06ac6bab59712cf78971be135c9 demoProject.jar
```

Ubuntu Linux Example:

```
$ sha256sum demoProject.jar  
65053d20874ec7929d219b24bdf98ac5b6f7b06ac6bab59712cf78971be135c9 demoProject.jar
```

At the time this document was written, there are no known means that would allow an attacker to produce a plugin module with different content, but the same SHA256 hash value of a particular file.

Terminology

Terminology

Algorithm instance - An algorithm instance is a fully-formed algorithm, which may be assigned to mask data in your masking Inventory. Algorithm instances are uniquely identified by their algorithmName in the Masking API, which is sometimes referred to as "algorithm code" or algorithmCd.

Algorithm component - This term refers to a Java class within an algorithm plugin that implements the MaskingAlgorithm Java interface.

Algorithm framework - This term refers to a family of algorithms on the Delphix Masking Engine. It is necessary to create an instance of an algorithm framework in order to use it - for example, FirstNameLookup is an instance of the Secure Lookup (aka. SL) algorithm framework.

Delphix algorithm SDK - A toolkit authored by Delphix to support the development of algorithm plugins. This includes a CLI for testing algorithms, a skeleton generator for creating empty plugin projects and algorithm classes, and sample algorithms illustrating various use cases.

Delphix masking API - This refers to the set of web APIs offered by the Delphix Masking Engine over HTTP/HTTPS. This API is sometimes referred to as the V5 APIs (referencing their current major version number) or Masking Web API.

Delphix masking plugin API - A package containing the set of Java interfaces that may be implemented in and consumed by a plugin for the Delphix Masking Engine. In order for a plugin to supply algorithms, one or more classes in the plugin must implement the MaskingAlgorithm interfaces provided by this API. This component also includes some common utilities used to load and run plugins on the engine and in the Masking SDK. The JAR containing the appropriate version of the Delphix Masking Plugin API classes has been embedded in the Algorithm SDK zip file.

Plugin - A JAR file containing classes that implement interfaces usable to extend the Delphix Masking Engine. Currently, only masking algorithms may be included in plugins. Plugins also contain self-descriptive metadata to facilitate their use on the engine.

Multi-Column (MC) algorithm - An algorithm that can take as input more than one field and mask one or all the inputted fields, computing the masked value using any of the fields provided. An MC Algorithm can also take in read-only fields that it does not modify but uses to compute a masked value for another field. The type of the input specified for an MC Algorithm is GENERIC_DATA_ROW, though all the fields must specify one of the "standard" masking types (STRING, BIG DECIMAL, etc).

Delphix masking APIs

This section covers the following topics:

- [Masking client](#)
- [API examples](#)

Masking client

This section covers the following topics:

- [Masking API client](#)
- [API calls for managing algorithms](#)
- [API calls for managing extended connectors](#)
- [API calls for ASDD profile set import and export](#)
- [API calls for managing classifiers](#)
- [API calls for searching and filtering](#)
- [API calls for managing masking job driver support tasks](#)
- [API calls for creating an inventory](#)
- [API calls for creating and running masking jobs](#)
- [API calls involving file upload and download](#)
- [Backwards compatibility API usage](#)
- [API response escaping](#)
- [API call for generating support bundle](#)

Masking API client

This section describes the API client available on the Masking Engine.

Introduction

With the release of API v5 on the Masking Engine, Delphix has opened up the possibility of scripting and automation against the Masking Engine. While this is exciting for us internally at Delphix, we are sure that this will be even more exciting for the consumers of the Masking Engine. This document is intended to be a high-level overview of what to expect with API v5 as well as some helpful links to get you started.

REST

API v5 is a RESTful API. REST stands for REpresentational State Transfer. A REST API will allow you to access and manipulate a textual representation of objects and resources using a predefined set of operations to accomplish various tasks.

JSON

API v5 uses JSON (JavaScript Object Notation) to ingest and return representations of the various objects used throughout various operations. JSON is a standard format and, as such, has many tools available to help with creating and parsing the request and response payloads, respectively.

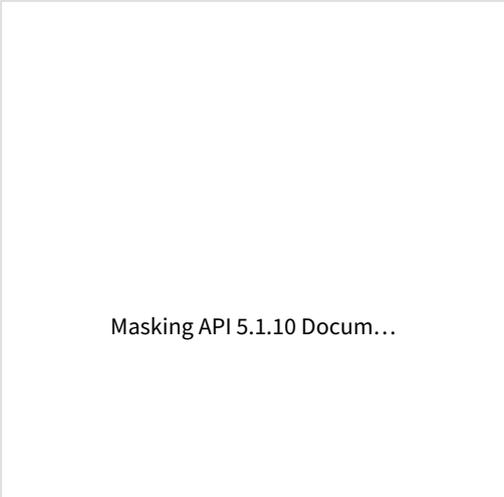
Here are some UNIX tools that can be used to parse JSON - <https://stackoverflow.com/questions/1955505/parsing-json-with-unix-tools>. That being said, this is only the tip of the iceberg when it comes to JSON parsing and the reader is encouraged to use their method of choice.

API client

The various operations and objects used to interact with API v5 are defined in a specification document. This allows us to utilize various tooling to ingest that specification to generate documentation and an API Client, which can be used to generate cURL commands for all operations. To see how to log into the API client and for some starter recipes, please check out API Cookbook document.

To access the API client on your Masking Engine, go to <http://myMaskingEngine.myDomain.com/masking/api-client>.

To access the API client documentation without an engine, please refer to the static HTML representations here:



Masking API 5.1.10 Docum...

Masking API 5.1.11 Docum...

Masking API 5.1.12 Docum...

Masking API 5.1.13 Docum...

Masking API 5.1.14 Docum...

Masking API 5.1.15 Docum...

Masking API 5.1.16 Docum...

Masking API 5.1.17 Docum...

Masking API 5.1.18 Docum...

Masking API 5.1.19 Docum...

Masking API 5.1.20 Docum...

Masking API 5.1.21 Docum...

Masking API 5.1.22 Docum...

Masking API 5.1.23 Docum...

Masking API 5.1.24 Docum...

Supported features

API v5 is in active development but does not currently support all features that are accessible in the GUI. The list of supported features will expand over the course of subsequent releases.

For a full list of supported APIs, the best place to look is the API client on your Masking Engine.

API calls for masking administration

The Delphix Masking Engine supports the following two types of administrative APIs:

- Analytics APIs
 - These APIs are for including Masking performance information in the support bundle and do not need to be used unless that information is requested.
- Application Setting APIs
 - Application Setting APIs allow an administrator to change the Delphix Masking Engine settings. Presently there are five categories of settings: analytics settings, LDAP settings, general settings, mask settings and profile settings. Over time, more settings will be added to give users direct control over the product's various settings. Below are the details of currently supported settings.

Application settings APIs

General group settings

Setting Group	Setting Name	Type	Description	Default Value
general	EnableMonitorRowCount	Boolean	Controls whether a job displays the total number of rows that are being masked. Setting this to false reduces the startup time of all jobs.	true
	PasswordTimeSpan	Integer [0, ∞)	The number of hours a user is locked out for before they can attempt to log in again.	23
	PasswordCount	Integer [0, ∞)	The number of incorrect password attempts before a user is locked out.	3
	AllowPasswordResetRequest	Boolean	When true, users can request a password reset link be sent to the email associated with their account.	true
	PasswordResetLinkDuration	Integer [1, ∞)	Controls how many minutes the password reset link is valid for.	5
	NumSimulJobsAllowed	Integer [0, ∞)	Max number of jobs allowed to run simultaneously. Setting this number to zero will lead to a dynamically calculated limit based on the number of available CPU cores.	7
	DefaultApiVersion	String	Used to set default API Version. If the version is omitted from the base path of the request's URL, but wishes to be treated using a specific masking API version that is not the latest version, set the DefaultApiVersion application setting. If the DefaultApiVersion is not set and the version is omitted from the URL, the latest version of the API on that engine will be used. Sample API Version format is like v5.1.5 etc.	Blank

Setting Group	Setting Name	Type	Description	Default Value
	DataRetentionInterval	Integer [-1, ∞)	The length of time that specific historical data is retained. This setting value is in integer days. Certain log files and internal processing data are retained in case problem diagnosis is needed. Since we cannot keep this data indefinitely, this setting is the length of time that old data is retained. Data older than this will be purged on a periodic basis. Special Values -1 : disable this pruning method 0 : each cleanup removes all files	60
	DataRetentionMaxDirectorySize	Integer [-1, 100]	The percentage of disk space allowed for all logfiles located in specific directories. This setting value is in integer percent. For log files written to disk, the DataRetentionInterval setting (above) ensures that we keep these job log files for only a specific period of time. This setting avoids problems where significant activity in a short time might overwhelm available disk space. This setting is a backstop to the DataRetentionInterval setting which is intended to be the primary driver for managing retention. Special Values -1 : disable this pruning method 0 : each cleanup removes all files	10

 NumSimulJobsAllowed setting should be set based on engine configuration. It is risky to run many jobs at once in an environment where you have scheduled more jobs than the system has memory for. When the system runs out of memory all jobs will fail.

Algorithm group settings

Setting Group	Setting Name	Type	Description	Default Value
algorithm	DefaultNonConformantDataHandling	String {DONT_MASK, FAIL}	Default algorithm behavior for Handling of Non-conformant Data patterns.	DONT_MASK

Database group settings

Setting Group	Setting Name	Type	Description	Default Value
database	DB2zDateFormat	String	Default Date String format to use for DB2 zOS if the database is not using one of the pre-defined IBM DB2 zOS Date String formats . Default is ISO Date String format.	yyyy-MM-dd

LDAP group settings

Setting Group	Setting Name	Type	Description	Default Value
ldap	Enable	Boolean	Used to enable and disable LDAP authentication	false
	LdapHost	String	Host of LDAP server	10.10.10.31
	LdapPort	Integer [0, ∞)	Port of LDAP server	389
	LdapBasedn	String	Base DN of LDAP server	DC=tbspune,DC=com
	LdapFilter	String	Filter for LDAP authentication	(&(objectClass=person) (sAMAccountName=?))
	MsadDomain	String	MSAD Domain for LDAP authentication	AD
	LdapTlsEnable	Boolean	Enable and disable the use of TLS for LDAP connections.	false

 In the LDAP group, once the "Enable" setting is set to "true", all users logging in will be authenticated via the LDAP server. Local authentication will no longer work. Before setting this to true set all other LDAP settings correctly and create the necessary LDAP users on the masking engine.

Mask group settings

Setting Group	Setting Name	Type	Description	Default Value
mask	DatabaseCommitSize	Integer [1, ∞)	Controls how many rows are updated (Batch Update) to the database before the transaction is committed.	10000
	DefaultStreams	Integer [1, ∞)	Default number of streams for a masking job.	1
	DefaultUpdateThreads	Integer [1, ∞)	Default number of database update threads for a masking job.	1
	DefaultMaxMemory	Integer [1024, ∞)	Default maximum memory for masking jobs (in megabytes).	1024
	DefaultMinMemory	Integer [1024, ∞)	Default minimum memory for masking jobs (in megabytes).	1024
	CharStreamingBufferLimitRate	Integer [1, 50]	Used for calculating maximum allowed buffer size for Character streaming parsers to buffer data. Only used in JSON file and Document store type masking.	25

Profile group settings

These settings apply only to the legacy profiler, not the ASDD profiler, unless specifically noted in the setting description.

Setting Group	Setting Name	Type	Description	Default Value
profile	EnableDataLevelCount	Boolean	<p>When enabled (true), the masking engine counts the number of rows in the profiled table. If the number of rows are less than or equal to DataLevelRows, then it uses the number of rows as the sample size. Otherwise, it uses DataLevelRows.</p> <p>When disabled (false), the masking engine uses DataLevelRows.</p>	false

Setting Group	Setting Name	Type	Description	Default Value
	DataLevelRows	Integer [1, ∞)	The number of rows a data level profiling job samples when profiling a column. The DataLevelRows will only take into account if <ul style="list-style-type: none"> • EnableDataLevelCount is false. • EnableDataLevelCount is true and number of rows is greater than DataLevelRows. 	100
	DataLevelPercentage	Double (0, ∞)	Percentage of rows that must match the data level regex to consider this column a match, and thus sensitive.	80.0
	IgnoreDatatype	String	Datatypes that a profiling job should ignore. Columns of these types will not be assigned a domain/algorithm pair.	BIT,BOOLEAN,CHAR#1,VARCHAR#1,VARCHAR2#1,NCHAR#1,NVARCHAR#1,NVARCHAR2#1,BINARY,VARBINARY,IMAGE,LOB,LONG,BLOB,CLOB,NCLOB,BFILE,RAW,ENUM,BFILE
	DefaultStreams	Integer [1, ∞)	Default number of streams for a profiling job.	1
	DefaultMaxMemory	Integer [1024, ∞)	Default maximum memory for profiling jobs (in megabytes).	1024
	DefaultMinMemory	Integer [1024, ∞)	Default minimum memory for profiling jobs (in megabytes).	1024

Setting Group	Setting Name	Type	Description	Default Value
	OptimizationLevel	Integer [0, 9)	Optimization level for the profiling job which is defined as below, 0: No optimizations are performed. 1: JavaScript runs in interpreted mode. 9: Performs the most optimization with faster script execution, but compiles slower. 1-9: All optimizations are performed.	-1
	DefaultMultiplePHIAlgorithm	String	Default Multiple PHI masking algorithm which will be used when the Multiple Profiler Expression will be true for profile job. This value is used by both the legacy and ASDD profilers.	NullValue Lookup

ASDD group settings

Setting Group	Setting Name	Type	Description	Default Value
ASDD	DefaultTableSampleRows	Integer [1, ∞)	The number of database rows for the ASDD profiler to sample for each table.	1000
	DefaultAssignmentThreshold	Integer [1, 100]	The confidence threshold that must be met or exceeded for the ASDD profiler to make a domain and algorithm assignment.	1
	DefaultJobExecutionStreams	Integer [1, ∞)	The number of streams to use by default for new ASDD profiler jobs	1
	DefaultNullFilterThreshold	Integer [0, 100]	The percentage of column values that must be null or empty to trigger an additional query to retrieve more column values.	75

Job group settings

Setting Group	Setting Name	Type	Description	Default Value
job	JobLoggingLevel	String {Basic, Detailed}	Controls the amount of information being logged from a job's output. Warning: the Detailed setting may log sensitive information when errors occur. Although this information can be very valuable when debugging a problem, it should be used with care.	Basic

CSP group CSP settings

Setting Group	Setting Name	Type	Description	Default Value
csp	CspFrameAncestorsDomain	String	Defines valid sources for embedding the resource using "frame", "iframe", "object", "embed", or "applet". To whitelist domains for frame-ancestors, add space-separated URLs in the CspFrameAncestorsDomain.	
csp	CspFormActionDomain	String	Defines valid sources that can be used as an HTML "form" action. To whitelist domains for form-action, add space-separated URLs in the CspFormActionDomain.	
csp	CspScriptSrcDomain	String	Defines valid sources of JavaScript. To whitelist domains for script-src, add space-separated URLs in the CspScriptSrcDomain.	
csp	StrictCspEnabled	Boolean	CSP setting should be enable via application setting based flag. This will provide more control to customer over CSP policy.	False

API calls for managing algorithms

This section covers the following topics:

- [Configuring algorithms](#)
- [Managing algorithm usage](#)
- [Migrating algorithms](#)
- [Binary lookup](#)
- [Character mapping](#)
- [Data cleansing](#)
- [Date replacement](#)
- [Date shift](#)
- [Dependent date shift](#)
- [Email](#)
- [Free text redaction](#)
- [Full name](#)
- [Mapping](#)
- [Min Max](#)
- [Name](#)
- [Numeric expression](#)
- [Payment card](#)
- [Regex decompose](#)
- [Secure lookup](#)
- [Segment mapping](#)
- [Tokenization](#)

Configuring algorithms

This section provides information on configuring algorithms using the API.

Masking client algorithm model

- **algorithmName** (maxLength=201)

String Equivalent to the algorithm name saved by the user through the GUI. For out of the box algorithms, this will be a similar name as that in the GUI, but presented in a more user-friendly format.

- **algorithmType**

String The type of algorithm Enum values: - BINARY_LOOKUP - CLEANSING - COMPONENT - LOOKUP - MAPPING - MINMAX - REDACTION - SEGMENT - TOKENIZATION

- **createdBy** (optional; readOnly; maxLength=255)

String The name of the user that created the algorithm

- **description** (optional; maxLength=255)

String The description of the algorithm

- **frameworkId**

Integer The frameworkId, corresponding to the framework that extensible algorithm is built upon. This field is to be used only for the Extensible Algorithms.

- **algorithmExtension** (optional)

Object Contains algorithm instance specific configuration parameters. See specific framework for more details.

Algorithm extension for extensible algorithms

It uses the generic Object, defined in the base AlgorithmExtension. Depending on the Extensible Algorithm design it currently supports following implementations (or their mix):

- **fileReference(s)** (optional, name is defined by Extensible Algorithm creator)

single fileReference or array[FileReference] A JSON formatted file reference or list of file references. Each file reference may be one of the following four options: - UUID value returned from the endpoint for uploading file to the Masking Engine - NFS mounted file URL - HTTP URL to external web located file - HTTPS URL to external web located file

- **calledExtendedAlgorithm(s)** (optional, name is defined by Extensible Algorithm creator)

single algorithmName or array[AlgorithmName] A JSON formatted name or a list of extensible algorithms names

Managing algorithm usage

Overview

The Masking Engine provides API endpoints to view and modify the usage of algorithms globally. These operations span all usage of an algorithm, including: database column and file format assignments across all environments, usage in domains, and references from other algorithms.

Viewing algorithm usage

The following API endpoint retrieves all usage of the algorithm specified in the request path:

```
algorithm GET algorithm/{algorithmName}/usage
```

This endpoint supports the following option in the query parameters:

- **includeAssignmentDetail** (optional, default=false)

boolean Enabling this option causes the API response to include a list of human-readable assignment detail objects, one for each individual usage of the algorithm in inventory. This can result in a very large response if the algorithm is heavily used. Algorithm usage in file formats will be reported once for each application of the file format to a file in inventory.

- **excludeChainedAlgorithms** (optional, default=false)

boolean Enabling this option causes the API response to exclude a list of algorithm references. This parameter will skip the computation of chained algorithms which will result in increased API performance.

- **environmentFilter** (optional)

String Report only usage occurring within the specified environment(s). This query option may be included multiple times, in which case usage for all matching environments will be reported. When the algorithm is used in a file format, all usage of that file format is reported so long as it is referenced by any environment matching the filter; this may include environments other than those selected by the filter. Filtering by environment excludes all domain and algorithm reference usage, as such usage is global rather than part of any particular environment.

- **rulesetFilter** (optional)

String Report only usage occurring within the specified ruleset(s). This query option may be included multiple times, in which case usage for all matching rulesets will be reported. When the algorithm is used in a file format, all usage of that file format is reported so long as it is referenced by any ruleset matching the filter; this may include rulesets other than those selected by the filter. Filtering by ruleset excludes all domain and algorithm reference usage, as such usage is global rather than part of any particular ruleset.

Updating algorithm usage

The following API endpoint updates all usage of the algorithm specified in the request path to use the new algorithm name supplied as a query parameter:

```
algorithm PUT algorithm/{algorithmName}/usage
```

This endpoint supports the following option in the query:

- **replacementAlgorithmName** (required, no default)

String The name of the algorithm that should replace the existing algorithm in all usage across the entire Masking Engine.

- **ignoreIncompatibleTypes** (optional, default=false)

boolean Setting this option to true will allow some algorithm replacements that would normally fail due to incompatible types to succeed. This may result in job failures if type conversions don't exist to convert the underlying data type to the type expected by the new algorithm.

- **environmentFilter** (optional)

String This options functions just like the environmentFilter option for the GET operation described above. Only usage matching the filter is updated. The update will fail if any file format referencing the algorithm is used from an environment that does not match the filter. Domain and algorithm reference usage is never updated when an environment filter is applied.

- **rulesetFilter** (optional)

String This options functions just like the rulesetFilter option for the GET operation described above. Only usage matching the filter is updated. The update will fail if any file format referencing the algorithm is used from a ruleset that does not match the filter. Domain and algorithm reference usage is never updated when a ruleset filter is applied.

The response body from the PUT request details all usage that was updated by the operation.

⚠ Globally updating algorithm usage can produce many inventory changes across multiple environments, and is not easily reversible when the replacement algorithm is already in use on the engine.

Delphix recommends performing the following steps *before* any update to algorithm usage via this API endpoint:

1. Perform the GET usage operation (described above) for both the existing and replacement algorithms. Carefully review the results and save them for future reference.
2. Export the engine's global settings and all affected environments prior to changing algorithm usage.

⚠ Algorithm compatibility checking of usage changes may still allow some replacements that could result in job failures using the new algorithm. Careful consideration should be given to whether the new algorithm can handle the data types and inputs for all usage of the algorithm being replaced.

Examples

Getting usage for an algorithm with one column assignment:

REQUEST

```
curl -X GET --header 'Accept: application/json' --header 'Authorization:
d46db68d-59f1-41e0-a128-c01bc920da30'
'http://masking-engine.example.com/masking/api/v5.1.10/algorithms/alg_6EBH8EGK/usage?
includeAssignmentDetail=false'
```

RESPONSE

```
{
  "algorithmName": "alg_6EBH8EGK",
  "columnMetadataIds": [
    11
  ],
  "fileFieldMetadataIds": [],
  "mainframeDatasetFieldMetadataIds": [],
  "domainNames": [
```

```

    "domain_6GXXQP60"
  ],
  "algorithmReferences": []
}

```

The same request with additional detail requested:

REQUEST

```

curl -X GET --header 'Accept: application/json' --header 'Authorization:
d46db68d-59f1-41e0-a128-c01bc920da30'
'http://masking-engine.example.com/masking/api/v5.1.10/algorithms/alg_6EBH8EGK/usage?
includeAssignmentDetail=true'

```

RESPONSE

```

{
  "algorithmName": "alg_6EBH8EGK",
  "columnMetadataIds": [
    11
  ],
  "fileFieldMetadataIds": [],
  "mainframeDatasetFieldMetadataIds": [],
  "domainNames": [
    "domain_6GXXQP60"
  ],
  "algorithmReferences": [],
  "assignmentDetails": [
    {
      "assignmentType": "DATABASE_COLUMN",
      "environmentName": "env_ZBQ0XK09",
      "databaseRulesetName": "rule_POQRBZ44",
      "databaseTableName": "profile",
      "databaseColumnName": "last_name"
    }
  ]
}

```

Updating all usage of algorithm named `alg_6EBH8EGK` to `alg_82U5GUZB` :

REQUEST

```

curl -X PUT --header 'Content-Type: application/json' --header 'Accept: application/
json'
--header 'Authorization: d46db68d-59f1-41e0-a128-c01bc920da30'
'http://masking-engine.example.com/masking/api/v5.1.10/algorithms/alg_6EBH8EGK/usage?
replacementAlgorithmName=alg_82U5GUZB&ignoreIncompatibleTypes=false'

```

RESPONSE

```

{

```

```
"columnMetadataIds": [  
  11  
],  
"fileFieldMetadataIds": [],  
"mainframeDatasetFieldMetadataIds": [],  
"domainNames": [  
  "domain_6G XKQP60"  
],  
"algorithmReferences": [],  
"assignmentDetails": []  
}
```

Migrating algorithms

Overview

As Delphix continues to make continuous improvement to the algorithms included with the Masking Engine, some algorithm frameworks will have multiple versions available simultaneously. New API paths have been added to allow migration of existing algorithm instances from old frameworks to new ones. The migration mechanism creates a new algorithm with the same configuration as the existing algorithm, allowing the behavior and performance of the migrated algorithm to be evaluated before adoption of the new algorithm for production use.

In this release, the following algorithm migrations are available:

- **FROM:** *algorithmType=MAPPING* **TO:** *algorithmType=COMPONENT, pluginName=dlpx-core, frameworkName=Mapping*

The [algorithm usage APIs](#) can be used to conveniently transition usage from the old to the new algorithm instance created by the migration mechanism.

Listing available migrations

The following API endpoint returns a list of result objects describing each possible migration. One object is returned for every algorithm on the engine that can be migrated:

```
algorithm GET algorithm/migration
```

Each object in the response contains the name of the algorithm that can be migrated, as well as the frameworkId of the framework that the migrated algorithm would use.

Migrating algorithms to new frameworks

The following API endpoint creates a new algorithm named **newAlgorithmName** (from the API query parameters), by migrating from the algorithm named in the query path:

```
algorithm POST /algorithms/{algorithmName}/migration
```

This endpoint requires the following option in the query:

- **newAlgorithmName** (required, no default)

String The name of the new algorithm to be created by the migration.

This response from the API is an AsyncTask object that can be used to check the status and result of the migration.



Migration of algorithms with a large amount of state (ex. a mapping algorithm with many mappings) can take several minutes or longer to complete. The engine's info.log will contain log messages indicating that the migration operation is making progress. Mapping algorithm migration is estimated to take approximately 3 minutes per 1,000,000 mapping values associated with the source algorithm.

Examples

Listing available migrations:

REQUEST

```
curl -X GET --header 'Accept: application/json' --header 'Authorization:
3d2d6f53-4b1a-42b5-b4c0-33ec3d66082f'
'http://masking-engine.example.com/masking/api/v5.1.10/algorithms/migration'
```

RESPONSE

```
{
  "availableMigrations": [
    {
      "algorithmName": "alg_J24QXMN3",
      "frameworkId": 13
    }
  ]
}
```

Migrating a mapping algorithm from the legacy framework to the new framework:

REQUEST

```
curl -X POST --header 'Content-Type: application/json' --header 'Accept: application/
json'
--header 'Authorization: 3d2d6f53-4b1a-42b5-b4c0-33ec3d66082f'
'http://masking-engine.example.com/masking/api/v5.1.10/algorithms/alg_J24QXMN3/
migration?newAlgorithmName=new_J24QXMN3'
```

RESPONSE

```
{
  "asyncTaskId": 29,
  "operation": "ALGORITHM_MIGRATE",
  "reference": "alg_J24QXMN3",
  "status": "WAITING",
  "cancellable": false
}
```

Binary lookup

See [Binary Lookup](#) for more information about this algorithm framework.

Creating a binary lookup algorithm via API

1. Retrieve the **frameworkId** for the BinarySL Framework. This information can be retrieved using the following endpoint:

```
algorithm GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 9,
  "frameworkName": "Binary Lookup",
  "frameworkType": "BYTE_BUFFER",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core",
    "pluginAuthor": "Delphix Engineering",
    "pluginType": "EXTENDED_ALGORITHM"
  }
}
```

2. Create a Binary SL algorithm instance via the following endpoint:

```
algorithm POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following:

```
{
  "algorithmName": "newbinarysl",
  "algorithmType": "COMPONENT",
  "description": "Binary Secure Lookup Example",
  "frameworkId": 9,
  "pluginId": 7,
  "algorithmExtension": {
    "lookupFiles": [
      {
        "uri": "delphix-file://upload/
f_39bbf352139f4234873109ad4e6271e1/file1.png"
      },
      {
        "uri": "delphix-file://upload/
f_093b5c07f90e4b9dbddb0339b71703d3/file2.png"
      },
      {

```

```
        "uri": "delphix-file://upload/  
f_8da2b97e201b4152b2befafc05612d8c/file3.png"  
      }  
    ]  
  }  
}
```

Binary SL algorithm extension

- **lookupFiles**(required, no default)

array[Object] A list of file reference UUID values returned from the endpoint for uploading files to the Masking Engine. There is a maximum limit of 50 files which can be uploaded into each instance of the algorithm

Character mapping

See [Character Mapping](#) for more information about this algorithm framework.

Creating a character mapping algorithm via API

1. Retrieve the **frameworkId** for the Character Mapping Framework. This information can be retrieved using the following endpoint:

```
algorithm GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 8,
  "frameworkName": "Chracter Mapping",
  "frameworkType": "STRING",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core"
  }
}
```

2. Create a Character Mapping algorithm instance via the following endpoint:

```
algorithm POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following:

```
{
  "algorithmName": "Digits and A to F",
  "algorithmType": "COMPONENT",
  "frameworkId": 8,
  "algorithmExtension": {
    "caseSensitive": true,
    "preserveRanges": null,
    "characterGroups": [
      "0123456789", "[a-fA-F]"
    ],
    "minMaskedPositions": 1,
    "preserveLeadingZeros": false
  }
}
```

Character mapping algorithm extension

- **characterGroups**(required, no default)

Array of Strings A list of String values defining the characters to be masked. Each group must be either: - a Java regex style character group beginning with '[' - a String of the literal characters that comprise the group. Duplication of characters within or among groups is not permitted.

- **caseSensitive**(default=true)

Boolean Whether the mapping should be case sensitive. When this is false, each group must be composed either: entirely of characters having no case; or of pairwise matching sets of LC and UC characters - example: [a-zA-Z], not [a-bC-D].

- **minMaskedPositions**(default=1, minimum=0)

Integer The minimum number of positions that must be replaced for masking to be considered successful. Non-conformant data handling is triggered whenever fewer positions are masked. Inputs containing only whitespace never trigger non-conformant data handling.

- **preserveRanges**(optional)

Array of PreserveRange objects A list of PreserveRange objects specifying regions of maskable characters to be preserved. Only maskable characters are considered when determining whether a position is preserved. Ranges are specified with position 0 representing the first maskable character.

- **preserveLeadingZeros**(default=false)

Boolean Whether to preserve leading '0' characters. This option may only be used when '0' is a masked character, and may not be used in conjunction with preserveRanges.

Character Mapping PreserveRange extension

- **start**(minimum=0, required, no default)

Integer The starting position of the preserve range, indexed starting with 0.

- **length**(minimum 1, required, no default)

Integer The length of the preserve range.

- **direction**(default="FORWARD")

String Defines the processing direction for this preserve range, with FORWARD starting at the beginning of input, and REVERSE starting at the end. Possible enum values: - FORWARD - process left to right - REVERSE - process right to left

Data cleansing

See [Data cleansing](#) for more information about this algorithm framework.

Creating a data cleansing algorithm via API

1. Retrieve the **frameworkId** for the Data Cleansing Framework. This can be done via the following endpoint:

```
algorithm GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 24,
  "frameworkName": "Data Cleansing",
  "frameworkType": "STRING",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core",
    "pluginAuthor": "Delphix Engineering",
    "pluginType": "EXTENDED_ALGORITHM"
  }
}
```

2. Upload a lookup file via the following endpoint:

```
fileUpload POST /file-uploads
```

Copy the **fileReferenceId** value returned in the Response Body.

3. Create a Data Cleansing algorithm via the following endpoint:

```
algorithm POST /algorithms
```

Using the [JSON formatted input](#), similar to the following example:

```
{
  "algorithmName": "demoDataCleansing",
  "algorithmType": "COMPONENT",
  "frameworkId": 24,
  "algorithmExtension": {
    "lookupFile": {
      "uri": "delphix-file://upload/f_52b19f8a9125435a83a1237fa53aeaf5/sample.txt"
    },
    "delimiter": "=",
    "caseSensitive": false,
    "trimWhitespace": true
  }
}
```

```
}
```

Data cleansing algorithm extension

- **lookupFile**(required)

String The `fileReferenceId` value returned from the `fileUpload` endpoint for uploading files to the Masking Engine. The file should contain a newline separated list of {value, replacement} pairs separated by the delimiter. No extraneous whitespace should be present.

- **delimiter**(required, minLength=1; maxLength=50; default="=")

String The delimiter string used to separate {value, replacement} pairs in the lookup file.

- **caseSensitive**(optional, default=true)

Boolean Whether the case of the input string must match the values in the lookup file.

- **trimWhitespace**(optional, default=true)

Boolean Whether to trim leading and trailing whitespace from the input string. Note: This must be **true** to cleanse fixed-width files and fixed-length database data types such as CHAR and NCHAR.

Date replacement

See [Date Replacement](#) for more information about this algorithm framework.

Creating a date replacement algorithm via API

1. Retrieve the **frameworkId** for the Date Replacement Framework. This information can be retrieved using the following endpoint:

```
algorithm GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 1,
  "frameworkName": "Date Replacement",
  "frameworkType": "LOCAL_DATE_TIME",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core"
  }
}
```

2. Create a Date Replacement algorithm via the following endpoint:

```
algorithm POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following:

```
{
  "algorithmName": "exampleDateReplacementAlgorithm",
  "algorithmType": "COMPONENT",
  "frameworkId": 1,
  "algorithmExtension": {
    "minDate": "2020-01-01 00:00:00",
    "maxDate": "2021-01-01 00:00:00",
    "unit": "DAYS"
  }
}
```

Date replacement algorithm extension

- **minDate**

| String A date representing the minimum value that an input can be masked to. The range is inclusive of this value.

- **maxDate**

| String A date representing the maximum value that an input can be masked to. The range is inclusive of this value.

- **unit**(default="SECONDS")

String A unit of time that determines what the output is truncated to. For example, when the unit is set to days, the years, months, and days may change, but the hours, minutes, and seconds will always be 00:00:00. Unit options supported by this framework: days, hours, minutes, and seconds.

Date shift

See [Date Shift](#) for more information about this algorithm framework.

Creating a date shift algorithm via API

1. Retrieve the **frameworkId** for the Date Shift Framework. This information can be retrieved using the following endpoint:

```
algorithm GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 5,
  "frameworkName": "Date Shift",
  "frameworkType": "LOCAL_DATE_TIME",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core"
  }
}
```

2. Create a Date Shift algorithm via the following endpoint:

```
algorithm POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following:

```
{
  "algorithmName": "exampleDateShiftAlgorithm",
  "algorithmType": "COMPONENT",
  "frameworkId": 5,
  "algorithmExtension": {
    "minRange": -3,
    "maxRange": 3,
    "unit": "MINUTES",
    "roll": "false"
  }
}
```

Date shift algorithm extension

- **minRange**

Integer A number representing the minimum range value from the input that the input can mask to. The range is inclusive of this value and must be an integer value. Negative values represent units of time in the past and positive

values represent units of time in the future. Zero may be included in the range or as one of the range values, but the input will not mask to the same value.

- **maxRange**

Integer A number representing the maximum range value from the input that the input can mask to. The range is inclusive of this value and must be an integer value. Negative values represent units of time in the past and positive values represent units of time in the future. Zero may be included in the range or as one of the range values, but the input will not mask to the same value.

- **unit(default="DAYS")**

String A unit of time that determines what the range is expressed in. Only one unit of time can be specified for each algorithm created. Masked values will be returned with the same granularity as the specified unit. For example a range of 1-2 days will not return the same masked values as a range of 24-48 hours as a range of 1-2 days will return a value with the hours, minutes, and seconds intact but a range of 24-48 hours may return a value with a change in hours anywhere from 24 hours to 48 hours. Unit options supported by this framework: years, months, days, hours, minutes, and seconds.

- **roll(default="false")**

String A boolean that represents whether or not the specified time unit should roll which means that units of time larger and smaller than the specified unit will remain the same. When set to false, there is no guarantee that larger units of time remain the same. When set to true, all larger units of time will retain their same values and the specified unit may wrap around to the beginning. For example, a date at the end of March may wrap around to the beginning of March while keeping all larger units of time and smaller units of time intact. Unit options supported by this framework: months, days, hours, minutes, and seconds.

Dependent date shift

See [Dependent Date Shift](#) for more information about this algorithm framework.

Creating a Dependent Date Shift algorithm via API

1. Retrieve the **frameworkId** for the Dependent Date Shift Framework. This information can be retrieved using the following endpoint:

```
algorithm GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 3,
  "frameworkName": "Dependent Date Shift",
  "frameworkType": "GENERIC_DATA_ROW",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core"
  }
}
```

2. Create a Dependent Date Shift algorithm via the following endpoint:

```
algorithm POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following:

```
{
  "algorithmName": "DependentDateShiftTest",
  "algorithmType": "COMPONENT",
  "createdBy": "admin",
  "description": "Test of the DependentDateShiftAlgo",
  "frameworkId": 3,
  "pluginId": 7,
  "fields": [
    {
      "fieldId": 1,
      "name": "date1",
      "type": "LOCAL_DATE_TIME"
    },
    {
      "fieldId": 2,
      "name": "date2",
      "type": "LOCAL_DATE_TIME"
    }
  ],
}
```

```

"algorithmExtension": {
  "roll": false,
  "unit": "DAYS",
  "maxRange": 5,
  "minRange": 3,
  "intervalRange": 2
}

```

Dependent Date Shift Algorithm extension

- **minRange**

Integer This number represents the smallest number of time units that will be added to date1 when masking. The range is inclusive of this value. Negative values represent units of time in the past and positive values represent units of time in the future. If date1 is not provided, this is applied to date2.

- **maxRange**

Integer This number represents the largest number of time units that will be added to date1 when masking. The range is inclusive of this value. Negative values represent units of time in the past and positive values represent units of time in the future. If date1 is not provided, this is applied to date2.

- **unit**(default="DAYS")

String A unit of time that the range is expressed in. This unit is also used to determine the interval between date1 and date2. Supported units include years, months, days, hours, minutes, and seconds.

- **roll**(default="false")

String A boolean that represents whether or not the specified time unit should roll which means that units of time larger and smaller than the specified unit will remain the same. When set to false, there is no guarantee that larger units of time remain the same. Option only supported for months, days, hours, minutes, and seconds. This applies when masking date1. If date1 is not provided, this is applied to date2

- **intervalRange**

Integer A number representing the +/- range value to shift the interval inclusive of the range value. A value of 0 will not change the interval between dates. This number may not be less than 0. If the specified unit difference between date1 and date2 is within the bound of the intervalRange, only values will be provided such that the sign of the difference is preserved. For example, if the day difference between date1 and date2 is 2 and the specified intervalRange is 3, only values greater than -2 will be used (i.e.: -1 to 3). Otherwise, the full range of values will be used (i.e.: -3 to 3).

Email

See [Email](#) for more information about this algorithm framework.

Creating an email algorithm via API

1. Retrieve the **frameworkId** for the Email Framework. This information can be retrieved using the following endpoint:

```
algorithm GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 3,
  "frameworkName": "Email",
  "frameworkType": "STRING",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core",
    "pluginAuthor": "Delphix Engineering",
    "pluginType": "EXTENDED_ALGORITHM"
  }
}
```

- a. Lookup files should be provided via File Reference. Files can be uploaded via the following endpoint: fileUpload POST /file-uploads Alternatively, those files might also be provided via HTTP / HTTPS / NFS mount URLs.
2. Create an Email algorithm via the following endpoint:

```
algorithm POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following:

```
{
  "algorithmName": "ExampleEmailAlgorithm",
  "algorithmType": "COMPONENT",
  "frameworkId": 3,
  "algorithmExtension": {
    "nameAction": "LOOKUP",
    "domainAction": "REPLACEMENT",
    "nameAlgorithm": null,
    "nameLookupFile": {
      "uri": "delphix-file://upload/f_08bb469a2ddc407bb97a31e96ed0a76a/lookup.txt"
    },
    "domainAlgorithm": null,
    "domainReplacementString": "delphix.com"
  }
}
```

}

Email algorithm extension

- **nameAction**

NameAction The type of action to apply to the name portion of the email. Must be one of the following enum values: - *UNIQUE* - applies a SHA-256 hash of the entire input then Base32 encodes the hash value - *LOOKUP* - applies a secure lookup using the values provided in the lookup list - *APPLY_ALGORITHM* - the name portion is replaced by the output of another chained masking algorithm

The **UNIQUE** option may produce masked name portions with lengths up to 52 characters.

- **domainAction**

MaskAction The type of action to apply to the name portion of the email. Must be one of the following enum values: - *REPLACEMENT* - the domain portion is replaced by a fixed value - *APPLY_ALGORITHM* - the domain portion is replaced by the output of another chained masking algorithm

- **nameLookupFile**

FileReference A file reference to a UTF-8 encoded file containing newline separated replacement values for the name portion of the email.

- **nameAlgorithm**

AlgorithmInstanceReference A reference for the algorithm to use when "APPLY_ALGORITHM" is the NameAction type. The algorithm must have maskingType "STRING". The algorithm will never be passed a null or empty value to mask. See *AlgorithmInstanceReference Extension* below for more information.

- **domainAlgorithm**

AlgorithmInstanceReference A reference for the algorithm to use when "APPLY_ALGORITHM" is the DomainAction type. The algorithm must have maskingType "STRING". The algorithm will never be passed a null or empty value to mask. See *AlgorithmInstanceReference Extension* below for more information.

- **domainReplacementString**

String

The string to replace the domain portion when "REPLACEMENT" is the DomainAction type.

AlgorithmInstanceReference Extension

- **name**

String The algorithm instance name.

Free text redaction

See [Free Text Redaction](#) for more information about this algorithm framework.

Creating a free text redaction algorithm via API

1. Retrieve the **frameworkId** for the Free Text Redaction Algorithm Framework. This information can be retrieved using the following endpoint:

```
algorithm GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 8,
  "frameworkName": "Free Text Redaction",
  "frameworkType": "STRING",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core",
    "pluginAuthor": "Delphix Engineering",
    "pluginType": "EXTENDED_ALGORITHM"
  }
}
```

2. Upload the **Lookup File** (if any) for the Free Text Redaction Algorithm Framework. It can be done using the following endpoint:

```
fileUpload POST /file-uploads
```

The response with the reference UUID information should look similar to the following:

```
{
  "fileReferenceId": "delphix-file://upload/f_6426ea480db14c1ea9f83f7eb98f3c0e/lookupFile.txt"
}
```

3. Create a Free Text Redaction Algorithm instance via the following endpoint:

```
algorithm POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following:

```
{
  "algorithmName": "Free Text Redaction for masking addresses and zip codes",
  "algorithmType": "COMPONENT",
  "frameworkId": 8,
}
```

```

    "algorithmExtension": {
      "isDenyList": true,
      "lookupFile": {
        "uri": "delphix-file://upload/
f_6426ea480db14c1ea9f83f7eb98f3c0e/lookupFile.txt"
      },
      "lookupFileRedactValue": "redact_value1",
      "regularExpressions": [
        {
          "patternString": "a|A"
        },
        {
          "patternString": "[0-9]{5}"
        }
      ],
      "regExRedactValue": "redact_value2"
    }
  }
}

```

Free text redaction algorithm extension

- **isDenyList** (required)

| *Boolean Deny list redaction if true, allow list redaction if false.*

- **lookupFile** (optional)

| *String The reference UUID value returned from the endpoint for uploading the lookup file to the Masking Engine.*

- **lookupFileRedactValue** (optional)

| *String The value to use to redact items matching entries specified in the lookup file.*

- **regExPatternList** (optional)

| *array[patternString]*

- **patternString**(required)

| *String Java 8 style regular expression.*

- **regExRedactValue** (optional)

| *String The value to use to redact items matching regular expression patterns.*

Full name

See [Full Name](#) for more information about this algorithm framework.

Creating a full name algorithm via API

1. Find the FrameworkId for the Extensible SL Framework. That might be done via the following EndPoint:

```
algorithm GET /algorithm/frameworks
```

Plugin name is **dlpx-core**, the framework name is **Full Name**.

2. Involved algorithm references might be built using the name of the desired existing extensible String-type algorithm. For example: "firstNameAlgorithmRef" : { "name" : "dlpx-core:FirstName" }
3. Create an Extensible Name Algorithm via the following EndPoint:

```
algorithm POST /algorithms
```

Using the [JSON formatted input](#), similar to the following example:

```
{
  "algorithmName": "demo-FullName",
  "algorithmType": "COMPONENT",
  "description": "This is a new style FullName algorithm",
  "frameworkId" : 3,
  "algorithmExtension" :
  {
    "firstNameAlgorithmRef" : { "name" : "dlpx-core:FirstName" },
    "lastNameAlgorithmRef" : { "name" : "dlpx-core:LastName" },
    "maxLengthOfMaskedName" : 0,
    "ifSingleWordConsiderAsLastName" : true,
    "lastNameAtTheEnd" : true,
    "lastNameSeparators" : [ "," ],
    "maxNumberFirstNames" : 2
  }
}
```

Fields description:

"algorithmName" - customer created algorithm name

"algorithmType" - should be "COMPONENT" for Extensible Algorithms

"description" - free text

"frameworkId" - the numeric value found in #1 above

"algorithmExtension" - the composite field, containing algorithm instance specific configuration parameters

Name algorithm extension

- **firstNameAlgorithmRef**(required)

| *AlgorithmReferenceId Must be an Algorithm Reference, pointing to an existing extensible algorithm of String type.*

- **lastNameAlgorithmRef**(required)

| *AlgorithmReferenceId* Must be an Algorithm Reference, pointing to an existing extensible algorithm of String type.

- **maxLengthOfMaskedName**(optional, default=0)

| *Integer* Should be a non-negative number. The output (masked) value is forcibly trimmed to that length (by the number of characters).

- **ifSingleWordConsiderAsLastName**(optional)

| *Boolean* If true consider single input word as a last name, otherwise as a first name. Default: true

- **lastNameAtTheEnd**(optional)

| *Boolean* If true last name to be detected at the end of the input string, otherwise last name is at the beginning. Default: true

- **lastNameSeparators**(optional)

| *List [Char]* List of the last name separators. Default: contains single value: comma ','

- **maxNumberFirstNames**(optional, default=2, minimum=1, maximum=4)

| *Integer* Defines the max number of first and middle names to be masked. The rest would be ignored.

Mapping

See [Mapping](#) for more information about this algorithm framework.

Creating a mapping algorithm via API

1. Retrieve the **frameworkId** for the Mapping Framework. This information can be retrieved using the following endpoint:

```
algorithm GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 15,
  "frameworkName": "Mapping",
  "frameworkType": "STRING",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core"
  }
}
```

2. Create a Mapping algorithm instance via the following endpoint:

```
algorithm POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following:

```
{
  "algorithmName": "MyMappingAlgo",
  "algorithmType": "COMPONENT",
  "frameworkId": 15,
  "algorithmExtension": {
    "ignoreCharacters": [],
    "mappingSet": {
      "host": "mypostgreshost.mydomain.com",
      "port": 5432,
      "schema": "mySchema",
      "database": "myDb",
      "isRemote": true,
      "algorithmName": "mappingTestRemote",
      "propertiesRef": {
        "uri": "delphix-file://upload/f_6ce20b134d5c4891bf90ccf7bd22d9b1/mapping.properties"
      }
    }
  }
}
```

```
}

```

i The above is an example of a remote mapping algorithm. See the extension options below for more information.

Mapping algorithm extension

- **ignoreCharacters** (optional; minimum=32; maximum=126)

array[Integer] The integer ASCII values of characters to ignore in the column data to map

- **mappingSet** (required)

mappingSet object An object that contains information about where the algorithm should find the mappings. See below for object property details.

MappingSet object

- **algorithmName** (required)

string The name of the algorithm this mappingSet corresponds to.

- **isRemote**

boolean Indicates if the mappings to be used for this algorithm are on the Masking Engine or if they are stored remotely. false if on the engine, true otherwise.

- **host**

string The host where the mapping database is running. Must be provided if isRemote is set to true.

- **port**

string The port to connect to the mapping database on the host. Must be provided if isRemote is set to true.

- **database**

string The name of the mapping database. Must be provided if isRemote is set to true.

- **schema**

string The schema where the mappings are. Must be provided if isRemote is set to true.

- **propertiesRef**

string The reference UUID value returned from the endpoint for uploading files to the Masking Engine. The file must be a properties file containing any further connection information for the database. Must be provided if isRemote is set to true.

Min Max

See [Min Max](#) for more information about this algorithm framework.

Creating a MinMax algorithm via API

1. Retrieve the **framework Id** for the Minmax Date or Minmax Number Algorithm Framework. This information can be retrieved using the following endpoint:

algorithm GET /algorithm/frameworks

The framework information should look similar to the following:

```
{
  "frameworkId": 8,
  "frameworkName": "MinMax Date",
  "frameworkType": "LOCAL_DATE_TIME",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core",
    "pluginAuthor": "Delphix Engineering",
    "pluginType": "EXTENDED_ALGORITHM"
  }
}
```

Or:

```
{
  "frameworkId": 15,
  "frameworkName": "MinMax Number",
  "frameworkType": "BIG_DECIMAL",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core",
    "pluginAuthor": "Delphix Engineering",
    "pluginType": "EXTENDED_ALGORITHM"
  }
}
```

2. Create a Free Text Redaction Algorithm instance via the following endpoint:

```
algorithm POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following (for MinMax Date algorithm framework):

```
{
  "algorithmName": "MinMax Date algorithm for 2021 year",
  "algorithmType": "COMPONENT",
  "frameworkId": 8,
  "algorithmExtension": {
```

```

        "minDate": "2021-01-01 00:00:00",
        "maxDate": "2021-12-31 00:00:00",
        "nonConformingDataDefaultValue": "replacement_value1"
    }
}

```

or the following (for MinMax Number algorithm framework):

```

{
    "algorithmName": "MinMax Number algorithm for normalizing age
range",
    "algorithmType": "COMPONENT",
    "frameworkId": 15,
    "algorithmExtension": {
        "minValue": 18,
        "maxValue": 65,
        "nonConformingDataDefaultValue": "replacement_value2"
    }
}

```

Minmax algorithm extension

- **minValue** (required)

Integer The minimum value for a Number range used in conjunction with *maxValue*. This field is used for "MinMax Number" framework only.

- **maxValue** (required)

Integer The maximum value for a Number range used in conjunction with and must be greater than *minValue*. This field is used for "MinMax Number" framework only.

- **minDate** (required)

date The minimum value for a Date range used in conjunction with *maxDate*. The Date must be specified in the following format: "yyyy-MM-dd HH:mm:ss". The Date will be interpreted as UTC. This field is used for "MinMax Date" framework only.

- **maxDate** (required)

date The maximum value for a Date range used in conjunction with and must be greater than *minDate*. The Date must be specified in the following format: "yyyy-MM-dd HH:mm:ss". The Date will be interpreted as UTC. This field is used for "MinMax Date" framework only.

- **nonConformingDataDefaultValue** (optional)

String The default replacement value for any value that is triggering non conforming data event handling. This field is only applicable when the underlying data to be masked is of type String and conversion to a Date or a Number is required.

Name

See [Name](#) for more information about this algorithm framework.

Creating a name algorithm via API

- Find the **frameworkId** for the Name Framework. This can be done via the following endpoint:

```
algorithm GET /algorithm/frameworks
```

The frameworkName is **"Name"** and the pluginName is **"dlpx-core"**.

```
{
  "frameworkId": 10,
  "frameworkName": "Name",
  "frameworkType": "STRING",
  "description": "This Name algorithm masks input name (first or last)
using values from the supplied LookupFile",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core",
    "pluginAuthor": "Delphix Engineering",
    "pluginType": "EXTENDED_ALGORITHM"
  }
}
```

- Involved files (lookupFile, particlesToRemoveFile, and particlesToPreserveFile) should be provided via the File Reference. They can be uploaded via the following endpoint:

```
fileUpload POST /file-uploads
```

Alternatively, those files can also be provided via HTTP / HTTPS / NFS mount URLs.

- Create an Extensible Name Algorithm via the following endpoint:

```
algorithm POST /algorithms
```

Using the [JSON formatted input](#), similar to the following example:

```
{
  "algorithmName": "NameDemo",
  "algorithmType": "COMPONENT",
  "description": "This is a new style Name algorithm",
  "frameworkId": 10,
  "algorithmExtension": {
    "lookupFile": {
      "uri": "delphix-file://upload/f_85f082535d054ee8a11696a24ed86d65/
LN_LOOKUP_100K.txt"
    },
    "particlesToRemoveFile": {
```

```

        "uri": "delphix-file://upload/f_1cc829ceee324113ab16c4e750dfce12/
particlesToRemove.txt"
    },
    "particlesToPreserveFile": {
        "uri": "delphix-file://upload/f_1cc829ceee324113ab16c4e750dfce12/
particlesToPreserve.txt"
    },
    "inputCaseSensitive": false,
    "filterAccent": true,
    "maskedValueCase": "PRESERVE_LOOKUP_FILE",
    "maxLengthOfMaskedName": 0,
    "maxNumberNames": 2
}
}
}

```

Fields description:

- **"algorithmName"**: User-defined algorithm name.
- **"algorithmType"**: Should be "COMPONENT" for Extensible Algorithms.
- **"description"**: User-defined, free text field.
- **"frameworkId"**: Numeric ID for the framework, provided in line #2 and #5 in the above excerpts.
- **"algorithmExtension"**: The composite field, containing algorithm instance specific configuration parameters.

Name algorithm extension

- **lookupFile** (required)

String Lookup file may be FileReferenceld in the one of the following four options: - UUID value returned from the endpoint for uploading file to the Masking Engine - NFS mounted file URL - HTTP URL to external web located file - HTTPS URL to external web located file

- **particlesToRemoveFile** (optional)

String File listing particles to remove may be FileReferenceld in the one of the following four options: - UUID value returned from the endpoint for uploading file to the Masking Engine - NFS mounted file URL - HTTP URL to external web located file - HTTPS URL to external web located file

- **particlesToPreserveFile** (optional)

String File listing particles to preserve may be FileReferenceld in the one of the following four options: - UUID value returned from the endpoint for uploading file to the Masking Engine - NFS mounted file URL - HTTP URL to external web located file - HTTPS URL to external web located file

- **inputCaseSensitive** (optional, default=false)

Boolean Setting "true" means input value case matter (i.e. "Peter" and "peter" might be masked to different values). Setting "false" (default) makes input value case insensitive ("Peter" and "peter" would be masked to the same value).

- **filterAccent** (optional, default=true)

Boolean

Setting "true" (default) means accented characters don't matter - similar input with and without accented characters are masked to the same values ("Adrián" and "Adrian" both mask to "John").

Setting "false" makes the input value accent sensitive ("Adrián" and "Adrian" would be masked to different values).

- **maskedValueCase** (optional, default="PRESERVE_INPUT")

String The output (masked) value case enforcing. Enum values: - PRESERVE_LOOKUP_FILE - use the unmodified replacement value.

- *PRESERVE_INPUT* - preserve case of input value. If mixed, use unmodified replacement value.
- *ALL_LOWER* - force the output to lowercase. - *ALL_UPPER* - force the output to uppercase.

- **maxLengthOfMaskedName** (optional, default=0)

Integer Should be a non-negative number. The output (masked) value is forcibly trimmed to that length (by the number of characters).

- **maxNumberNames** (optional, default=2, minimum=1, maximum=4)

Integer

Defines the max number of names to be masked and returned. The rest are removed.

Numeric expression

See [Numeric Expression](#) for more information about this algorithm framework.

Creating a numeric expression algorithm via API

1. Retrieve the **frameworkId** for the Numeric Expression Framework. This information can be retrieved using the following endpoint:

```
algorithm    GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 26,
  "frameworkName": "Numeric Expression",
  "frameworkType": "BIG_DECIMAL",
  "description": "Numeric Expression masks input by ... [truncated for
brevity].",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core",
    "pluginAuthor": "Delphix Engineering",
    "pluginType": "EXTENDED_ALGORITHM"
  }
}
```

2. Create a Numeric Expression algorithm instance via the following endpoint:

```
algorithm    POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following:

```
{
  "algorithmName": "NumericExpressionTest",
  "algorithmType": "COMPONENT",
  "frameworkId": 26,
  "algorithmExtension": {
    "expression": "Math.floor(((input * randomPercentage) * 100.0) + 0.5) /
100.0",
    "inputType": "DOUBLE",
    "constants": [
      {
        "name": "randomPercentage",
        "value": "new java.util.Random(seed).doubles(0.1,
0.9).iterator().nextDouble()"
      }
    ],
    "nonConformingDataDefaultValue": "100.0"
  }
}
```

```
}
}
```

Numeric expression algorithm extension

- **expression**

String One-line mathematical expression written in the Java programming language that references `input` (the current unmasked value), e.g. `input * 0.5` or `input + Math.random()`.

- **inputType**

String ENUM(DOUBLE, LONG, BIG_DECIMAL) Data type that `input` conforms to within the expression. `DOUBLE` (default) is double-precision floating point, `LONG` is long integer, and `BIG_DECIMAL` is `java.math.BigDecimal` object.

- **constants**(optional)

array[Constant] An array of `Constant` objects. Constants are variables that the expression can reference by name and whose values remain fixed for the life of a masking job. Constants can reference by name other constants defined before them.

- **nonConformingDataDefaultValue**(optional)

String Default masked value to be used if the unmasked input is not a numeric data type and can't automatically be converted to one.

Constant

- **name**

String Must be valid Java variable name. No two constants can have the same name, nor can "input" or "seed" be used as a constant name.

- **value**

String One-line Java expression that must return a value, which is not required to be numeric.

Payment card

See [Payment Card](#) for more information about this algorithm framework.

Creating a payment card algorithm via API

1. Retrieve the **frameworkId** for the Payment Framework. This information can be retrieved using the following endpoint:

```
algorithm GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 4,
  "frameworkName": "Payment Card",
  "frameworkType": "STRING",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core"
  }
}
```

2. Create a Payment Card algorithm via the following endpoint:

```
algorithm POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following:

```
{
  "algorithmName": "examplePaymentCardAlgorithm",
  "algorithmType": "COMPONENT",
  "frameworkId": 4,
  "algorithmExtension": {
    "minMaskedPositions": 7,
    "preserve": 4
  }
}
```

Payment card algorithm extension

- **minMaskedPositions**(default=1, minValue=0, maxValue=32)

Integer A value that represents the minimum number of positions that must be replaced for masking to be considered successful. A non-conformant data error is thrown when fewer positions are masked. The minimum value for this field is 0 and the default value is 1. The maximum value is 32.

- **preserve**(default=0, minValue=0, maxValue=32)

Integer A value that represents the number of maskable characters to preserve at the beginning of the input. Only maskable characters are considered when determining whether a position is preserved. The minimum value for this field is 0 and the default value is 0. The maximum value is 32.

Regex decompose

See [Regex Decompose](#) for more information about this algorithm framework.

Creating a regex decompose algorithm via API

1. Retrieve the **frameworkId** for the Regex Decompose Framework. This information can be retrieved using the following endpoint:

```
algorithm GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 5,
  "frameworkName": "Regex Decompose",
  "frameworkType": "STRING",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core",
    "pluginAuthor": "Delphix Engineering",
    "pluginType": "EXTENDED_ALGORITHM"
  }
}
```

2. Create a Regex Decompose algorithm via the following endpoint:

```
algorithm POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following:

```
{
  "algorithmName": "ExampleRegexDecomposeAlgorithm",
  "algorithmType": "COMPONENT",
  "frameworkId": 5,
  "algorithmExtension": {
    "trimInput": true,
    "requireMask": false,
    "maskPatterns": [
      {
        "regex": "([0-9]+)-([a-z]+)",
        "actions": [
          {
            "type": "REDACT",
            "algorithm": null,
            "redactString": "asdf",
            "redactCharacter": null
          },
          {

```

```

        "type": "REDACT",
        "algorithm": null,
        "redactString": null,
        "redactCharacter": "x"
      }
    ]
  },
  "fallbackAction": null,
  "maxLength": 65536
}

```

Regex decompose algorithm extension

- **maskPatterns**

Array of MaskPattern objects Defines the mask pattern(s) of the algorithm. See [Regex Decompose MaskPattern Extension](#) below for more information.

- **fallbackAction**

MaskAction The action that should be applied to the entire input if none of the defined regular expressions match. If no pattern matches and no fallbackAction is set, non-conformant data handling will be triggered. See [Regex Decompose MaskAction Extension](#) below for more information.

- **requireMask** (default="true")

String A boolean that represents whether the input must be masked. When this is true, patterns are matched until one changes the input. If no pattern can change the input and no fallbackAction is set, non-conformant data handling will be triggered for this value. If false, the first matching pattern will apply regardless of whether it changes the input. Any difference in value from the input is considered successful masking.

- **trimInput** (default="true")

String A boolean that represents whether to trim whitespace from the beginning and end of the input prior to processing. The same leading and trailing whitespace will be reintroduced into the masked value. This option is provided to simplify the regular expressions that can be used in maskPatterns, as they no longer must account for and preserve leading and trailing whitespace.

- **maxLength** (default=65536, minValue=1)

Integer A value that represents the maximum character length of input the algorithm will attempt to process. If the input length exceeds this value, non-conformant data handling will be triggered for this value.

Regex decompose maskpattern extension

- **regex**

String A Java 8 style regular expression used to match the masking input.

- **actions**

Array of MaskAction objects Defines the action(s) to be applied to the match or capturing group(s) when the regular expression matches. See [Regex Decompose MaskAction Extension](#) below for more information.

Regex decompose mask action extension

- **type**

String The type of action to the input that matches the regex. Must be one of the following enum values: - PRESERVE - the value or capturing group is not masked and remains unchanged - TRUNCATE - the value or capturing group is

replaced with "" - REDACT - the value or capturing group is replaced by a value or repeated character -
APPLY_ALGORITHM - the value or capturing group is replaced by the output of another chained masking algorithm

- **algorithm**

AlgorithmInstanceReference A reference for the algorithm to use when "APPLY_ALGORITHM" is the MaskAction type. The algorithm must have maskingType "STRING". The algorithm will never be passed a null or empty value to mask. See *AlgorithmInstanceReference Extension* below for more information.

- **redactCharacter**

String The character to use to replace the input when "REDACT" is the MaskAction type. Each character in the portion of input matched is replaced with this character. Length of the matched input is preserved. Only one of *redactCharacter* or *redactString* may be specified for a given MaskAction.

- **redactString**

String The string to use to replace the input when "REDACT" is the MaskAction type. The entire matched portion is replaced with this string. Use of this option will cause the length of the value to change during masking unless the matched portion of input happens to have the same length of the *redactString*. Only one of *redactCharacter* or *redactString* may be specified for a given MaskAction.

AlgorithmInstanceReference Extension

- **name**

String The algorithm instance name.

Secure lookup

See [Secure Lookup](#) for more information about this algorithm framework.

Creating a secure lookup algorithm via API

1. Find the frameworkId for the Extensible SL Framework. This can be done via the following endpoint:

```
algorithm GET /algorithm/frameworks
```

Plugin name is **dlpx-core**, the framework name is **Secure Lookup**.

2. Upload Lookup File via the following endpoint:

```
fileUpload POST /file-uploads
```

Alternatively, the Lookup File might also be provided via HTTP / HTTPS / NFS mount URLs.

3. Create an Extensible SL Algorithm via the following endpoint:

```
algorithm POST /algorithms
```

Using the [JSON formatted input](#), similar to the following example:

```
{
  "algorithmName": "demoExtendedSL",
  "algorithmType": "COMPONENT",
  "frameworkId" : 1,
  "algorithmExtension" :
  {
    "lookupFile" : {
      "uri":"delphix-file://upload/f_7984ee9672b44e309f7ef5940f856e7c/
ColorsLF.txt"
    },
    "inputCaseSensitive" : true,
    "maskedValueCase" : "ALL_LOWER",
    "hashMethod" : "SHA256"
  }
}
```

Fields description:

- "algorithmName": customer created algorithm name
- "algorithmType": should be "COMPONENT" for Extensible Algorithms
- "description": free text
- "frameworkId": the numeric value found in #1 above
- "algorithmExtension": the composite field, containing algorithm instance specific configuration parameters

Exporting secure lookup values via API

Secure lookup values can now be exported from algorithms. These values can only be exported from algorithms of type **LOOKUP** or type **COMPONENT** where the framework name is **Secure Lookup**.

1. Find the algorithmCd of the algorithm instance to retrieve the values from. This may be done via the following endpoint:

```
algorithm GET /algorithms
```

2. Use the following endpoint to export the lookup values:

```
algorithm POST /algorithms/{algorithmName}/export-lookup-values
```

Info:

Lookup values cannot be exported from algorithms where the lookup values are provided via MOUNT or via HTTP/HTTPS.

3. A response similar to the following will be returned:

```
{
  "asyncTaskId": 55,
  "operation": "EXPORT_SL_VALUES",
  "reference": "EXPORT_SL_VALUES-c2VjdXJlbG9va3VwX2NNSGdZc2FQLnR4dA==",
  "status": "WAITING",
  "cancellable": false
}
```

4. Retrieve the "reference" from the response body in the previous step and use this value as the fileDownloadId for the following endpoint:

```
fileDownload GET /file-downloads/{fileDownloadId}
```

The response will contain the exported lookup values. Values will be returned in a plaintext file with newline-separated values.

Secure Lookup algorithm extension

- **lookupFile** (maxLength=255)

String Lookup file may be one of the following four options: - UUID value returned from the endpoint for uploading file to the Masking Engine - NFS mounted file URL - HTTP URL to external web located file - HTTPS URL to external web located file

- **inputCaseSensitive** (optional, default=false)

Boolean Setting "true" means input value case matter (i.e. "Peter" and "peter" might be masked to different values) Setting "false" (default) makes input value case-insensitive ("Peter" and "peter" would be masked to the same value)

- **maskedValueCase** (optional, default="PRESERVE_LOOKUP_FILE")

String The output (masked) value case enforcing. Enum values: - PRESERVE_LOOKUP_FILE - use the unmodified replacement value (default). - PRESERVE_INPUT - preserve case of input value. If mixed - use unmodified replacement value. - ALL_LOWER - force the output to lowercase. - ALL_UPPER - force the output to uppercase.

- **hashMethod** (optional, default="SHA256")

String The hash method used to select replacement values. Must be one of the following enum values: - SHA256 - the default hash method for extensible secure lookup - LEGACY - hash method used to mimic the legacy secure lookup behavior in the extensibility framework

Segment mapping

See [Segment Mapping](#) for more information about this algorithm framework.

Creating a segment mapping algorithm via API

1. Retrieve the **frameworkId** for the Segment Mapping Framework. This information can be retrieved using the following endpoint:

```
algorithm    GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 6,
  "frameworkName": "Segment Mapping",
  "frameworkType": "STRING",
  "description": "The Segment Mapping Algorithm will ... [truncated for brevity].",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core",
    "pluginAuthor": "Delphix Engineering",
    "pluginType": "EXTENDED_ALGORITHM"
  }
}
```

2. Create a Segment Mapping algorithm instance via the following endpoint:

```
algorithm    POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following:

```
{
  "algorithmName": "SegmentMappingTest",
  "algorithmType": "COMPONENT",
  "frameworkId": 6,
  "algorithmExtension": {
    "segments": [
      {
        "length": 4,
        "segmentType": "MASK_NUMERIC",
        "inputValues": null,
        "maskValues": null
      },
      {
        "length": 1,
        "segmentType": "PRESERVE"
      }
    ]
  }
}
```

```

        {
            "length": 2,
            "segmentType": "MASK_ALPHANUMERIC",
            "inputValues": "A,B,C,F-G",
            "maskValues": "Q-S,X,Y,Z"
        },
        "ignoreCharacters": [],
        "autoIgnoreCharacters": true,
        "allowShortSegments": false,
        "processPreserveBeforeIgnore": false
    }
}

```

Segment mapping algorithm extension

- **segments** (required, minimum=1, maximum=10)

Array of Segment objects A list of Segment Mapping Segments defining the masking behavior in order. See [Segment Mapping Segment Extension](#) below for more information.

- **ignoreCharacters** (optional)

Array of Integers A list of integer ASCII values of characters to ignore. For example, [44, 65] would ignore commas and the letter 'A'. These are removed from input value before masking and restored to their original positions after masking.

- **autoIgnoreCharacters** (default=false)

Boolean Whether or not to ignore all non alpha-numeric characters. Use this as an alternative to specifying individual characters in ignoreCharacters.

- **allowShortSegments** (default=false)

Boolean Whether or not to allow masking of short MASK_NUMERIC segments. When set to false, a MASK_NUMERIC segment cannot be masked if it is shorter than the defined segment length. If a short MASK_NUMERIC segment is encountered, a NonConformantDataException will be triggered. When set to true, a short MASK_NUMERIC segment may be masked. **Note:** If set to true and a MASK_NUMERIC segment is defined, the algorithm is not reversible and cannot be used for tokenization/re-identification.

- **processPreserveBeforeIgnore** (default=false)

Boolean Whether or not to process PRESERVE segments before removing ignore characters. When set to false, ignore characters are removed from the input string first, and then all segments are processed in the order in which they are defined. When set to true, PRESERVE segment are processed first, before removing ignore characters, so the preserved segment positions are based on the original input string, which may include ignore characters. Afterwards, ignore characters are removed and the remaining string is masked according to the other segment definitions. **Setting this to true is not recommended, as it may cause some segments to be processed out of order.**

Segment mapping segment extension

- **length** (required, minimum=1, maximum=6)

Integer The length of the segment in characters.

- **segmentType** (required)

String The masking behavior for this segment. Enum values: - MASK_ALPHANUMERIC - mask letters to letters and digits to digits. Mappings are configured for each character position independently (e.g. AA -> GC, 'A' does not always mask to the same letter at each position) - MASK_NUMERIC - mask the entire segment as a single integer

value to another integer value - PRESERVE - do not mask this segment - CONSTANT - mask any input value to a constant value

- **inputValues**

String Defines the input values to mask in this segment, provided as either individual values, ranges, or a combination thereof. This field is optional for MASK_ALPHANUMERIC and MASK_NUMERIC, and if left blank or omitted (null), the default value ranges are used. This field is not used for PRESERVE or CONSTANT. For a MASK_ALPHANUMERIC segment, the default value range is '0-9,A-Z'. You can specify something like 'A-F,P,R,1-5,7,9'. For a MASK_NUMERIC segment, the default value range is 0 to the max integer that can fit into the segment length (ex: 000-999 for a segment of length 3). You can specify integer values and ranges, like '10,30,50-875'. The masking will only look to mask these values and will preserve any other values.

- **maskValues**

String Defines the values to mask to for this segment. This is defined the same way as inputValues and has the same default value ranges. This field is optional for MASK_ALPHANUMERIC and MASK_NUMERIC, and if left blank or omitted (null), the default value ranges are used. This field is required for CONSTANT and not used for PRESERVE.

Note: if the inputValues and maskValues are not the same, then the algorithm is not reversible and cannot be used for tokenization/re-identification.

Tokenization

See [Tokenization](#) for more information about this algorithm framework.

Creating a tokenization algorithm via API

1. Retrieve the **frameworkId** for the Tokenization Framework. This information can be retrieved using the following endpoint:

```
algorithm GET /algorithm/frameworks
```

The framework information should look similar to the following:

```
{
  "frameworkId": 13,
  "frameworkName": "Tokenization",
  "frameworkType": "STRING",
  "plugin": {
    "pluginId": 7,
    "pluginName": "dlpx-core",
    "pluginAuthor": "Delphix Engineering",
    "pluginType": "EXTENDED_ALGORITHM"
  }
}
```

2. Create a Tokenization algorithm instance via the following endpoint:

```
algorithm POST /algorithms
```

Configure a new algorithm using the [JSON formatted input](#) similar to the following:

```
{
  "algorithmName": "exampleTokenization",
  "algorithmType": "COMPONENT",
  "frameworkId": 13,
  "algorithmExtension": {
    "ivLength": 16,
    "fallback": "CHARACTER_MAPPING",
    "cmCharacterGroups": [
      "[A-Za-z0-9+/"
    ],
    "cmMinMaskedPositions": 1
  }
}
```

Tokenization algorithm extension

- **ivLength**(default=16, minimum=0, maximum=16)

Integer The length of the initialization vector (IV) used for AES in CBC-CTS mode. The default length is 16, which offers the most security. The tradeoff is that this increases the length of the masked result. Selecting a lower IV length decreases the length of the masked result. It is recommended that you only select an IV length of 0 if you

require the masked value for each input to be consistent between jobs and for the same input to only mask to one output.

- **fallback**(required, no default)

String This specifies how to handle masking a value where the encrypted result does not fit in the column size. If an AES encrypted result is too long to fit into the field, there are two fallback options: - NONE - the job fails if the masked result is too long - CHARACTER_MAPPING - the Character Mapping algorithm is used to tokenize the value, which produces a result that is the same length as the input

Extension for Character Mapping fallback

- **cmCharacterGroups**(default=["[A-Za-z0-9+/]"])

Array of Strings A list of String values defining the characters to be masked. Each group must be either: - a Java regex style character group beginning with '[' - a String of the literal characters that comprise the group. Duplication of characters within or among groups is not permitted.

- **cmMinMaskedPositions**(default=1, minimum=0)

Integer The minimum number of positions that must be replaced for masking to be considered successful. Non-conformant data handling is triggered whenever fewer positions are masked. Inputs containing only whitespace never trigger non-conformant data handling.

API calls for managing extended connectors

Introduction

This section details how to manage extended database connectors, including how to manage driver support tasks on a masking job.

1. [Installing a driver support plugin](#)
2. [Installing a JDBC driver](#)
3. [Creating an extended database connector](#)
4. [Managing masking job driver support tasks](#)



Installing a JDBC driver with a driver support is only possible via the web API.

Installing a driver support plugin

Install driver support jar on masking engine

1. Select `POST /file-uploads`
2. Click "Choose File" and select desired driver support jar

The response will look similar to the following with a return status of 200:

```
{
  "fileReferenceId": "delphix-file://upload/f_xxxx/sampleDriverSupport.jar"
}
```

Create driver support plugin

1. Select `POST /plugins`
2. **fileReferenceId:** delphix-file://upload/f_xxxx/sampleDriverSupport.jar
3. **pluginName:** whatever desired name
4. **pluginType:** DRIVER_SUPPORT

The response will look similar to the following with a return status of 200:

```
{
  "pluginId": 9,
  "pluginName": "Sample Plugin",
  "pluginAuthor": "Sample Plugin Author",
  "pluginType": "DRIVER_SUPPORT",
  "originalFileName": "driverSupport.jar",
  "originalFileChecksum":
"f8398c0768ecf7709c6992b3f048f9da8be640285b3ccc968973949ca3cceb02",
  "installDate": "2021-04-21T15:29:01.982+00:00",
  "installUser": 5,
  "builtIn": false,
  "pluginVersion": "1.5.0",
  "pluginObjects": [
    {
      "objectIdentifier": "1",
      "objectName": "Disable Constraints",
      "objectType": "DRIVER_SUPPORT_TASK"
    },
    {
      "objectIdentifier": "2",
      "objectName": "Disable Triggers",
      "objectType": "DRIVER_SUPPORT_TASK"
    },
    {
      "objectIdentifier": "3",
      "objectName": "Drop Indexes",
      "objectType": "DRIVER_SUPPORT_TASK"
    }
  ]
}
```

```
    ]
  }
```

i The `objectIdentifier` field refers to the ID of the task. Specifying the ID of the tasks is required to [enable/disable tasks](#) on a masking job. `objectIdentifier` (task ID) has no bearing on the task execution order. The task order is determined by the order the tasks are added to `getTasks` in the [Driver Support Plugin implementation](#).

Create JDBC driver that uses driver support plugin

1. Select `POST /jdbc-drivers` (or `PUT /jdbc-drivers/{jdbcDriverId}` to update existing JDBC driver)
2. Form the request body as follows:

```
{
  "driverName": "HANA driver",
  "driverClassName": "com.sap.db.jdbc.Driver",
  "fileReferenceId": "delphix-file://upload/f_xxxx/sampleJdbcDriver.zip",
  "driverSupportId": 9
}
```

The response will look similar to the following with a return status of 200:

```
{
  "jdbcDriverId": 8,
  "driverName": "HANA driver",
  "driverClassName": "com.sap.db.jdbc.Driver",
  "version": "2.4",
  "uploadedBy": "admin",
  "uploadDate": "2021-04-27T20:34:47.748+00:00",
  "checksum": "a5b7cf1323b71398e68fd583cd4f40ef8a5f4212ae94b63e95c904ed226d4c7b",
  "builtIn": false,
  "loggerInstalled": true,
  "driverSupportId": 9
}
```

⚠ If the referenced driver support plugin is being used by existing masking jobs that have tasks enabled, extra validation is performed. In the case of updating a driver support plugin or updating a JDBC driver to use a different driver support, the driver support plugin must implement all enabled tasks on any existing masking job. If the other driver support does not implement all enabled tasks, the update will fail. In the case of deleting a driver support plugin, the delete will fail if the driver support plugin is being used by any existing masking jobs that have tasks enabled.

Installing a JDBC driver

Install JDBC driver zip on masking engine

1. Select `POST /file-uploads`
2. Click "Choose File" and select desired JDBC driver zip

The response will look similar to the following with a return status of 200:

```
{
  "fileReferenceId": "delphix-file://upload/f_xxxx/sampleJdbcDriver.zip"
}
```

 Note that you can also install a JDBC driver [via the UI](#).

Create JDBC driver without driver support

1. Select `POST /jdbc-drivers`
2. Format the request body as follows:

```
{
  "driverName": "HANA driver",
  "driverClassName": "com.sap.db.jdbc.Driver",
  "fileReferenceId": "delphix-file://upload/f_xxxx/sampleJdbcDriver.zip",
}
```

The response will look similar to the following with a return status of 200:

```
{
  "jdbcDriverId": 8,
  "driverName": "HANA driver",
  "driverClassName": "com.sap.db.jdbc.Driver",
  "version": "2.4",
  "uploadedBy": "admin",
  "uploadDate": "2021-04-27T20:34:47.748+00:00",
  "checksum": "a5b7cf1323b71398e68fd583cd4f40ef8a5f4212ae94b63e95c904ed226d4c7b",
  "builtIn": false,
  "loggerInstalled": true,
}
```

Create JDBC driver with driver support

To create a JDBC driver with driver support, follow the same process, but add `driverSupportId` to the request body. This is used to specify the ID of the driver support plugin to associate with the JDBC driver.

 If the JDBC driver's referenced driver support plugin tasks are enabled on any existing masking job, validation on update is done in order to prevent changing the driver support plugin to another one unless it implements all enabled tasks. If the other driver support does not implement all enabled tasks, the update will fail.

Creating an extended database connector

Creating an extended database connector

 This assumes an application and environment already exists, to which you can add this extended connector.

1. Select `POST /database-connectors`
2. Format response body as follows:

```
{
  "connectorName": "hana db",
  "databaseType": "EXTENDED",
  "environmentId": 1,
  "jdbc": "JDBC_SERVER_URL",
  "username": "USERNAME",
  "password": "PASSWORD",
  "kerberosAuth": false,
  "jdbcDriverId": 7,
  "enableLogger": false
}
```

The response will look similar to the following with a return status of 200:

```
{
  "databaseConnectorId": 1,
  "connectorName": "hana db",
  "databaseType": "EXTENDED",
  "environmentId": 1,
  "jdbc": "JDBC_SERVER_URL",
  "username": "USERNAME",
  "kerberosAuth": false,
  "jdbcDriverId": 7,
  "enableLogger": false
}
```

Managing masking job driver support tasks

For information on managing masking driver support tasks, see [API Calls for Managing Masking Job Driver Support Tasks](#).

API calls for ASDD profile set import and export

A profile set defines the set of classifiers or expressions that will be used to identify sensitive information in the rule set when a profiling job is run. Refer to [Discovering your sensitive data](#) for an overview of profile sets and related concepts. More information about ASDD profile set import and export can be found in the [ASDD profile set import and export article](#). More information about configuring profile sets can be found in the [Configuring Profile Sets article](#).

In order to perform ASDD profile set import and export using the API client, you should have some familiarity with REST APIs and JSON data encoding. Access the API Client, as described in the [Masking API Client](#) section, and authenticate by pressing the **Authorize** button at the upper right part of the screen.

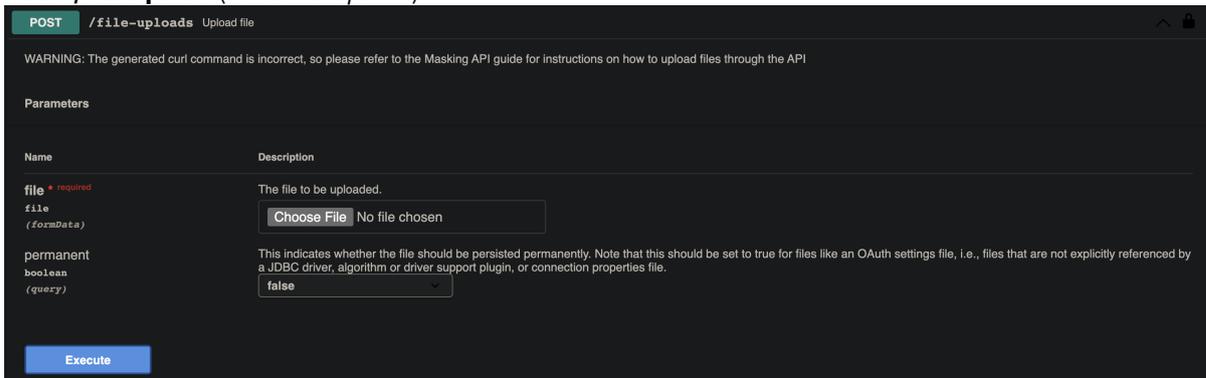
The ASDD profile set import and export endpoints utilize the file upload and download endpoints, respectively.

 The generated curl command for file upload and download visible in the API client is incorrect. If using curl over the API client is desired, see [API calls involving file upload and download](#) for more information on how to format file upload and download curl requests correctly.

ASDD profile set import

Import requires two API calls:

1. **POST /file-upload** (under *fileUpload*)



POST /file-uploads Upload file

WARNING: The generated curl command is incorrect, so please refer to the Masking API guide for instructions on how to upload files through the API

Parameters

Name	Description
file * required file (formData)	The file to be uploaded. <input type="button" value="Choose File"/> No file chosen
permanent boolean (query)	This indicates whether the file should be persisted permanently. Note that this should be set to true for files like an OAuth settings file, i.e., files that are not explicitly referenced by a JDBC driver, algorithm or driver support plugin, or connection properties file. <input type="checkbox"/> false

2. **POST /configuration-install** (under *configuration*)

POST /configuration-install Create installation

Parameters

Name	Description
body * required object (body)	The configuration to install

Edit Value Model

```
{
  "fileReferenceId": "delphix-file://upload/f_08bb469a2ddc407bb97a31e96ed0a76a/foo.bar"
}
```

Cancel

Parameter content type
application/json

Execute

These API path's purposes are as expected based on the path and operation; the purpose of **POST /file-upload** is to upload the profile set configuration zip file to the engine file system, and the purpose of **POST /configuration-install** is to unpack the zip file and import the profile set configuration onto the engine.

Example

In this example, a new profile set is configured that has a new domain and a few classifiers (one PATH classifier, one LIST and one TYPE). The content of the Standard.json file is as follows:

```
{
  "profileSetName": "Custom ASDD Profile Set",
  "domains": [
    {
      "domain": "CUSTOM_PII",
      "maskingAlgorithmName": "dlpx-core:CM Alpha-Numeric"
    }
  ],
  "classifiers": [
    {
      "domain": "CUSTOM_PII",
      "name": "Capital - Path",
      "type": "PATH",
      "properties": {
        "paths": [
          {
            "allowPartialMatch": true,
            "caseSensitive": false,
            "fieldValue": "(?i)capital",
            "matchStrength": 0.67,
            "matchType": "REGEX",
            "parentValue": ""
          }
        ]
      }
    },
    {
```

```

        "allowPartialMatch": true,
        "caseSensitive": false,
        "fieldValue": "(?i)capital[ _-]??(city)?",
        "matchStrength": 0.5,
        "matchType": "REGEX",
        "parentValue": ""
    }
],
    "rejectStrength": 0.0
}
},
{
    "domain": "CUSTOM_PII",
    "name": "Capital - Type",
    "type": "TYPE",
    "properties": {
        "allowedTypes": [
            {
                "minimumLength": 10,
                "typeName": "String"
            }
        ],
        "matchAutoIncrementingColumn": false
    }
},
{
    "domain": "CUSTOM_PII",
    "name": "Capitals - List",
    "type": "LIST",
    "properties": {
        "rejectStrength": 0.5,
        "valueLists": [
            {
                "file": "file://de_capitals.txt",
                "matchStrength": 1.0
            }
        ]
    }
}
]
}
}

```

The above profile set configuration file (Standard.json) and the file used by the LIST classifier (de_capitals.txt) are contained in a zip file that is uploaded to the engine filesystem via **POST /file-upload** (with the default permanent query parameter value of *false*). The API response is:

```

{
  "fileReferenceId": "delphix-file://upload/f_8f90c13fb1b9434abdbe871da894a2a8/
custom_profile_set_1.zip",
  "filename": "custom_profile_set_1.zip",
  "fileSize": 3724,
  "persistenceType": "OBJECT/TEMPORARY"
}

```

```
}

```

The `fileReferenceId` value returned by **POST /file-upload** is used for the **POST /configuration-install** request to initiate profile set import:

```
{
  "fileReferenceId": "delphix-file://upload/f_8f90c13fb1b9434abdbe871da894a2a8/custom_profile_set_1.zip"
}
```

The API response is:

```
{
  "fileReferenceId": "delphix-file://upload/f_8f90c13fb1b9434abdbe871da894a2a8/custom_profile_set_1.zip",
  "requiresRestart": false
}
```

ASDD profile set export

Export requires two API calls:

1. **POST /profile-sets/{profileSetId}/export**
2. **GET /file-downloads/{fileDownloadId}**

This API path's purpose is as expected based on the path and operation; the purpose of **POST /profile-sets/{profileSetId}/export** is to export the given profile set's configuration in a zip file in the same format expected for import. This API call is asynchronous, so it spawns an async task with a file reference that you can then download via **GET /file-downloads/{fileDownloadId}** when the task is complete. The API response for the custom profile set imported above (**POST /profile-sets/5/export**) is:

```
{
  "asyncTaskId": 1,
  "operation": "EXPORT_PROFILE_SET",
  "reference": "EXPORT_PROFILE_SET-cHJvZm1sZV9zZXRFNS56aXA=",
  "status": "WAITING",
  "cancellable": false
}
```

GET /file-downloads/EXPORT_PROFILE_SET-cHJvZm1sZV9zZXRFNS56aXA= then provides a link to download the file in the response body.

API calls for managing classifiers

Classifier instances define logic used by the ASDD profiler to identify sensitive information. Refer to [Discovering your sensitive data](#) for an overview of classifiers and related concepts. Each classifier instance is based on a **classifier framework** that implements the recognition logic. An overview of the available frameworks is available under the [classifier concept section](#). More information about managing classifiers can be found in the [Managing classifiers article](#).

i Creating or modifying profile sets can also be done via the API and requires only two API calls See [ASDD profile set import and export](#) for usage instructions.

In order to manage classifiers using the API client, you should have some familiarity with REST APIs and JSON data encoding. Access the API Client, as described in the [Masking API client](#) section, and authenticate by pressing the **Authorize** button at the upper right part of the screen. Locate the API paths related to classifiers:

classifier			^
GET	/classifiers	Get all classifiers	v 🔒
POST	/classifiers	Create classifier	v 🔒
GET	/classifiers/{classifierId}	Get classifier by ID	v 🔒
PUT	/classifiers/{classifierId}	Update classifier by ID	v 🔒
DELETE	/classifiers/{classifierId}	Delete classifier by ID	v 🔒
POST	/classifiers/{classifierId}/copy	Copy classifier by ID	v 🔒
POST	/classifiers/{classifierId}/export-files	Export files used for specified classifier	v 🔒
GET	/classifiers/frameworks	Get all classifier frameworks	v 🔒
GET	/classifiers/frameworks/{frameworkId}	Get classifier framework by frameworkid	v 🔒
POST	/classifiers/search	Search classifiers	v 🔒

Each of these API path's purposes are as expected based on the operation; GET to view the configuration of existing classifiers, PUT to modify the configuration of an existing classifier, POST to create a new classifier, and DELETE to delete a classifier.

Retrieving classifier framework configurations

The **classifiers/frameworks** paths allow retrieval of information about the available classifier frameworks. In particular, making a request to these endpoints with `include_schema=true` will return open API style descriptions of the schema for the classifier frameworks. It is also necessary to use this API to map classifier framework type names, such as 'PATH' or 'REGEX', to the numeric frameworkId when creating classifier instances.

When creating a new classifier, it can be helpful to first perform a GET operation that retrieves the configuration of an existing classifier instance, using the intended framework as a starting point.

Example: Creating a new PATH classifier

In this example, a database contains some columns named **snack_pref1**, **snack_pref2**, etc. The columns contain sensitive user data that should be masked, thus, a good regex for recognizing these columns would be **snack_pref[0-9]+**. A **SNACK_PREF** domain has been created with an appropriate algorithm for this type of data. To match the column name to profile, the type of classifier needed is **PATH**.

First, perform a GET operation on the **classifiers/frameworks** path. In the output below, the respective **frameworkId** of the **PATH** classifier is **3**.

```

{
  "_pageInfo": {
    "numberOnPage": 4,
    "total": 4
  },
  "responseList": [
    {
      "frameworkId": 1,
      "frameworkName": "REGEX",
      "description": "The regex framework can be used to specify one or more regular
expressions to match the data in a field."
    },
    {
      "frameworkId": 2,
      "frameworkName": "LIST",
      "description": "The list framework can be used to specify one or more value
lists to match the data in a field."
    },
    {
      "frameworkId": 3,
      "frameworkName": "PATH",
      "description": "The path framework can be used to specify exact values or
regular expressions to match the name of a field."
    },
    {
      "frameworkId": 4,
      "frameworkName": "TYPE",
      "description": "The type framework can be used to specify valid types and type
lengths for fields to rule out invalid data types and lengths during classification."
    }
  ]
}

```

After determining that **classifier 1** has **frameworkId=3**, perform the GET operation on **classifiers/1**:

```

{
  "classifierId": 1,
  "classifierName": "Account Number - Path",
  "frameworkId": 3,
  "domainName": "ACCOUNT_NO",
  "createdBy": "System",
  "builtIn": true,
  "classifierConfiguration": {
    "paths": [
      {
        "matchType": "REGEX",
        "fieldValue": "(?i)(?>(account|acct|acnt)_?-? ?(number|num|nbr|no|user))$",
        "parentValue": "",
        "caseSensitive": false,
        "matchStrength": 0.67,
        "allowPartialMatch": true
      }
    ]
  }
}

```

```

    ],
    "rejectStrength": 0
  }
}

```

To exemplify, this configuration is edited, replacing several configuration values to create a new classifier:

```

{
  "classifierName": "Snack Preference - Path",
  "frameworkId": 3,
  "domainName": "SNACK_PREF",
  "classifierConfiguration": {
    "paths": [
      {
        "matchType": "REGEX",
        "fieldValue": "snack_pref[0-9]+",
        "parentValue": "",
        "caseSensitive": false,
        "matchStrength": 0.67,
        "allowPartialMatch": false
      }
    ],
    "rejectStrength": 0
  }
}

```

This body can then be used with a POST operation to the **classifiers** path to create the new classifier. The API response will include the newly assigned **classifierId**.

LIST type classifiers

Note that LIST type classifiers require one or more input files to define the value lists for recognition. These files must first be uploaded by doing a POST to the **fileUpload API endpoint**. The resulting **fileReferenceId** values may then be used for fields of type FILE in the classifier configuration when creating the classifier.

Downloading files associated with classifiers

Using the **classifiers/{classifierId}/export-files** endpoint, files associated with classifiers can be downloaded by specifying the **classifierId**. This endpoint will return an async task similar to the following:

```

{
  "asyncTaskId": 2,
  "operation": "EXPORT_CLASSIFIER_FILES",
  "reference": "EXPORT_CLASSIFIER_FILES-Y2xhc3NpZml1c180NV92YWx1ZV9saXN0cy56aXA=",
  "status": "RUNNING",
  "startTime": "2023-03-21T18:45:59.390+00:00",
  "cancellable": false
}

```

You can check the status of this task by using the

asyncTask/{asyncTaskId} endpoint.

```
{
  "asyncTaskId": 2,
  "operation": "EXPORT_CLASSIFIER_FILES",
  "reference": "EXPORT_CLASSIFIER_FILES-Y2xhc3NpZml1c180NV92YWx1ZV9saXN0cy56aXA=",
  "status": "SUCCEEDED",
  "startTime": "2023-03-21T18:45:59.390+00:00",
  "endTime": "2023-03-21T18:45:59.520+00:00",
  "cancellable": false
}
```

Once this task has succeeded, the **fileDownload/{fileDownloadId}** endpoint can be used to download the files where **fileDownloadId** is the **reference** provided in the **classifiers/{classifierId}/export-files** API response. The files are returned in a .zip that contains all files associated with the specified classifier.

Searching and Filtering Classifiers

The **classifiers/search** endpoint allows for searching and filtering of classifiers. More information on syntax can be found at [API calls for searching and filtering](#).

API calls for searching and filtering

API endpoints are available for certain objects that allow searching and filtering. The body of these endpoints takes a single key-value pair `{“filter_expression“: “<expression>”}` that should conform to the filter-query language specified below.

The following objects support searching and filtering:

- [classifiers](#)
- [profileExpressions](#)
- [profileTypeExpressions](#)

Comparison operators

Operator	Description	Example
CONTAINS	Substring or membership testing for string and list attributes respectively.	field3 CONTAINS 'foobar', field4 CONTAINS TRUE
IN	Tests if field is a member of a list literal. List can contain a maximum of 100 values.	field2 IN ['Goku', 'Vegeta']
GE	Tests if a field is greater than or equal to a literal value	field1 GE 1.2e-2
GT	Tests if a field is greater than a literal value	field1 GT 1.2e-2
LE	Tests if a field is less than or equal to a literal value	field1 LE 9000
LT	Tests if a field is less than a literal value	field1 LT 9.02
NE	Tests if a field is not equal to a literal value	field1 NE 42
EQ	Tests if a field is equal to a literal value	field1 EQ 42

Search operator

The SEARCH operator filters for items that have any filterable attribute that contains the input string as a substring. The comparison is done case-insensitively. This is not restricted to attributes with string values. Specifically ``SEARCH '12'`` would match an item with an attribute with an integer value of ``123``.

Logical operators

Logical operators are ordered by precedence.

Operator	Description	Example
NOT	Logical NOT (right associative)	NOT filed1 LE 9000
AND	Logical AND (left associative)	field1 GT 9000 AND field2 EQ 'Goku'
OR	Logical OR (left associative)	field1 GT 9000 OR field2 EQ 'Goku'

Grouping

Parentheses '()' can be used to override operator precedence.

For example:

- NOT (field1 LT 1234 AND field2 CONTAINS 'foo')

Literal Values

Literal	Description	Example
Nil	Represents the absence of a value	nil, Nil, nll, NIL
Boolean	true/false boolean	true, false, True, False, TRUE, FALSE
Number	Signed integer and floating point numbers. Also supports scientific notation.	0, 1, -1, 1.2, 0.35, 1.2e-2, -1.2e+2
String	Single or double quoted	"foo", "bar", "foo bar", 'foo', 'bar', 'foo bar'
Datetime	Formatted according to RFC3339	2018-04-27T18:39:26.397237+00:00
List	Comma-separated literals wrapped in square brackets	[0], [0, 1], ['foo', "bar"]

Limitations

- Not all fields of the objects are filterable or searchable. The allowed fields for the specific object are listed in the body description of that object API endpoint.
- For a filterable field of the form-*field1[field2]*, the query within **filter_expression** should start with the "**contains**" operator.
 - For example, for a filterable-field of "domain[domain_name]" the query body should be **{“filter_expression”: "domain contains {domain_name EQ ‘Account_NO’}”}**
- A maximum of 8 unique identifiers may be used inside a filter expression.

Filtering usage examples

Below is a sample 'fruit_inventory' table containing information that will be filtered using the above syntax.

id	name	color	size	quantity	in_season
1	apple	red	medium	4	false
2	watermelon	red	large	1	true
3	strawberry	red	small	10	true
4	orange	orange	medium	7	false
5	kiwi	green	small	3	false
6	raspberry	red	small	20	false
7	lemon	yellow	medium	2	true
8	lime	green	small	8	false
9	pineapple	yellow	large	3	true
10	blueberry	blue	small	132	true

Example 1:

This example uses the CONTAINS operator to search for the substring "berry" in the name field.

```
{"filter_expression": "name CONTAINS 'berry'"}
```

This query returns the following objects:

id	name	color	size	quantity	in_season
3	strawberry	red	small	10	true
6	raspberry	red	small	20	false
10	blueberry	blue	small	132	true

Example 2:

This example uses the GT operator to filter for quantities greater than 5 and the EQ operator to filter for size "small".

```
{"filter_expression": "quantity GT 5 AND size EQ 'small'"}
```

This query returns the following objects:

id	name	color	size	quantity	in_season
3	strawberry	red	small	10	true
6	raspberry	red	small	20	false
8	lime	green	small	8	false
10	blueberry	blue	small	132	true

Example 3:

This example uses the NOT operator to filter for objects where the color is not in list ['red', 'orange', 'green'].

```
{"filter_expression": "NOT color IN ['red','orange','green']"}
```

This query returns the following objects:

id	name	color	size	quantity	in_season
7	lemon	yellow	medium	2	true
9	pineapple	yellow	large	3	true
10	blueberry	blue	small	132	true

Example 4:

This example uses the EQ operator to filter for objects where the boolean value of in_season is equal to true.

```
{"filter_expression": "in_season EQ true"}
```

This query returns the following objects:

id	name	color	size	quantity	in_season
2	watermelon	red	large	1	true

id	name	color	size	quantity	in_season
3	strawberry	red	small	10	true
7	lemon	yellow	medium	2	true
9	pineapple	yellow	large	3	true
10	blueberry	blue	small	132	true

Example 5:

This example uses a combination of many operators with grouping to filter for objects that are green, small, and have a quantity greater than or equal to 8, or objects that are medium, not in season, and are in list ['apple', 'lemon'].

```
{"filter_expression": "(color EQ 'green' AND size EQ 'small' AND quantity GE 8) OR (size EQ 'medium' AND in_season EQ false AND name IN ['apple', 'lemon'])"}
```

This query returns the following objects:

id	name	color	size	quantity	in_season
1	apple	red	medium	4	false
8	lime	green	small	8	false

Example 6:

This example demonstrates the filter query on an 'order' table (described below), which is joined to the 'fruit_inventory' table (above) by the fruit-'name' column. Use the 'contains' query only when the allowed filterable_attribute in the API mentions

explicitly like 'order[order_id]' instead of 'order_id'.

order_id	name	order_quantity
1	apple	2
2	strawberry	5
3	lime	7

```
{"filter_expression": "order contains {name EQ 'lime'}"}
```

This query returns the following objects:

order_id	name	order_quantity
3	lime	7

API calls for managing masking job driver support tasks

Enabling driver support tasks is possible for built-in Oracle and MSSQL connectors as well as extended connectors that [have a JDBC driver that uses a driver support plugin](#) at the following endpoints:

- Masking jobs - POST /masking-jobs and PUT /masking-jobs/{maskingJobId}
- Reidentification jobs - POST /reidentification-jobs and PUT /reidentification-jobs/{reidentificationJobId}
- Tokenization jobs - POST /tokenization-jobs and PUT /tokenization-jobs/{tokenizationJobId}

Disabling driver support tasks is possible for built-in Oracle and MSSQL connectors as well as extended connectors that [have a JDBC driver that uses a driver support plugin](#) at the following endpoints:

- PUT /masking-jobs/{maskingJobId}
- PUT /reidentification-jobs/{reidentificationJobId}
- PUT /tokenization-jobs/{tokenizationJobId}

i The order of the tasks returned in `enabledTasks` in the Job APIs' responses is not indicative of the task execution order. The task order is determined by the order the tasks are added to `getTasks` in the [Driver support plugin implementation](#).

The following instructions to enable driver support tasks on an Oracle masking job can be used to enable driver support tasks for applicable reidentification and tokenization jobs as well.

View the tasks implemented by driver support plugin

1. Select GET /plugin (or GET /plugin/{pluginId} if the plugin ID of the driver support is known).
2. Change `pluginType` query parameter to `DRIVER_SUPPORT` (default is `EXTENDED_ALGORITHM`).
3. The response should include the full list of driver support plugins on the masking engine. If the engine only has the builtin Oracle driver support plugin installed, the response will look as follows:

```
{
  "_pageInfo": {
    "numberOnPage": 1,
    "total": 1
  },
  "responseList": [
    {
      "pluginId": 8,
      "pluginName": "dlpx-oracle-driver-support",
      "pluginAuthor": "Delphix Engineering",
      "pluginType": "DRIVER_SUPPORT",
      "originalFileName": "delphix-oracle-driver-support-plugin-1.0.0.jar",
      "originalFileChecksum":
"17b06f2fd888888e26a634d501b4ac9be5a91a7f50000a995934145c7afe7e12",
    }
  ]
}
```

```

    "installDate": "2021-10-24T18:08:50.868+00:00",
    "builtIn": true,
    "pluginVersion": "1.0.0",
    "description": "This plugin provides built-in driver support
functionality for the Oracle JDBC driver that ships with the Delphix Masking
Engine.",
    "pluginObjects": [
      {
        "objectIdentifier": "1",
        "objectName": "Disable Constraints",
        "objectType": "DRIVER_SUPPORT_TASK"
      },
      {
        "objectIdentifier": "2",
        "objectName": "Drop Indexes",
        "objectType": "DRIVER_SUPPORT_TASK"
      },
      {
        "objectIdentifier": "3",
        "objectName": "Disable Triggers",
        "objectType": "DRIVER_SUPPORT_TASK"
      }
    ]
  }
]
}

```

Create masking Job that enables tasks

i This assumes a ruleset using the desired connector already exists. The following example demonstrates the creation of an in-place masking job on a built-in Oracle connector. This also assumes you know the ID of the task that you want to enable and have execute as part of a given masking job. To enable tasks to execute as part of a masking job on an *extended* connector, you need to ensure the ruleset points to an extended connector that is using a JDBC driver with a driver support and include the property `enabledTasks` in your request.

1. Select `POST /masking-jobs` to create a masking job using the ruleset you created earlier that targets the desired connector.
2. Format the request body as follows to enable Disable Constraints, Drop Indexes and Disable Triggers per the `objectIdentifier` values returned from the GET Plugin API endpoint:

```

{
  "jobName": "Oracle IP job",
  "rulesetId": 1,
  "jobDescription": "Job description",
  "enabledTasks": [
    {
      "taskId": 1
    }
  ],
}

```

```

    {
      "taskId": 2
    },
    {
      "taskId": 3
    }
  ]
}

```

The response will look similar to the following with a return status of 200:

```

{
  "maskingJobId": 1,
  "jobName": "Oracle IP job",
  "rulesetId": 1,
  "rulesetType": "table",
  "createdBy": "admin",
  "createdTime": "2021-04-27T21:29:46.043+00:00",
  "feedbackSize": 50000,
  "jobDescription": "Job description",
  "maxMemory": 1024,
  "minMemory": 1024,
  "multiTenant": false,
  "numInputStreams": 1,
  "onTheFlyMasking": false,
  "databaseMaskingOptions": {
    "batchUpdate": true,
    "commitSize": 10000,
    "disableConstraints": false,
    "dropIndexes": false,
    "disableTriggers": false,
    "numOutputThreadsPerStream": 1,
    "truncateTables": false
  },
  "failImmediately": false,
  "enabledTasks": [
    {
      "taskId": 1
    },
    {
      "taskId": 2
    },
    {
      "taskId": 3
    }
  ],
  "streamRowLimit": 20000
}

```

Disable tasks

To disable the Disable Triggers task on an Oracle masking job, the request body to `PUT /masking-jobs/1` should exclude the taskId of the task to disable. Using the above request body as an example, Disable Triggers has a task ID of 3 so the request body to `PUT /masking-job/1` should exclude the object in `enabledTasks` with `"taskId": 3`. The request body should thus be:

```
{
  "jobName": "Oracle IP job",
  "rulesetId": 1,
  "jobDescription": "Job description",
  "onTheFlyMasking": false,
  "enabledTasks": [
    {
      "taskId": 1
    },
    {
      "taskId": 2
    }
  ]
}
```

The Oracle masking job will now only have Disable Constraints and Drop Indexes enabled (in this example, their respective task IDs are 1 and 2). The response will look similar to the following with a return status of 200:

```
{
  "maskingJobId": 1,
  "jobName": "Oracle IP job",
  "rulesetId": 1,
  "rulesetType": "table",
  "createdBy": "admin",
  "createdTime": "2021-04-27T21:29:46.043+00:00",
  "feedbackSize": 50000,
  "jobDescription": "Job description",
  "maxMemory": 1024,
  "minMemory": 1024,
  "multiTenant": false,
  "numInputStreams": 1,
  "onTheFlyMasking": false,
  "databaseMaskingOptions": {
    "batchUpdate": true,
    "commitSize": 10000,
    "disableConstraints": false,
    "dropIndexes": false,
    "disableTriggers": false,
    "numOutputThreadsPerStream": 1,
    "truncateTables": false
  },
  "failImmediately": false,
  "enabledTasks": [
    {
```

```
        "taskId": 1
    },
    {
        "taskId": 2
    }
],
"streamRowLimit": 20000
}
```

API calls for creating an inventory

Below are examples of requests you might enter and responses you might receive from the Masking API client. For commands specific to your masking engine, work with your interactive client at <http://<myMaskingEngine>/masking/api-client/>

 HTTPS (SSL/TLS) is recommended, but for explanatory purposes these examples use insecure HTTP

 In all code examples, replace `<>` with the hostname or IP address of your virtual machine.

Fetch table names from database connector

Object references you will need:

- The ID of the database connector to fetch tables for

 This database connector ID (1, in this example) is included in the PATH for this operation, NOT the payload.

REQUEST

```
curl -X GET --header 'Accept: application/json' --header 'Authorization:
7c856e3d-5b20-4261-b5fe-cc2ffcee5ae0'
'http://<myMaskingEngine>/masking/api/database-connectors/1/fetch'
```

RESPONSE

```
[ "ALL_COLUMNS", "DBVERIFICATION_TABLE" ]
```

More info

```
http://<myMaskingEngine>/masking/api-client/#!/databaseConnector/fetchTableMetadata
```

Example

See how to use this in the context of a script [here](#).

Create table metadata

Object references you will need:

- The name of the table to create the metadata for
- The ruleset ID

REQUEST

```
curl -X POST --header 'Content-Type: application/json' --header 'Accept:
application/json' --header 'Authorization:
7c856e3d-5b20-4261-b5fe-cc2ffcee5ae0' -d '{ "tableName": "ALL_COLUMNS",
"rulesetId": 2 }'
'http://<myMaskingEngine>/masking/api/table-metadata'
```

RESPONSE

```
{ "tableMetadataId": 2, "tableName": "ALL_COLUMNS", "rulesetId": 2
}
```

More info

<http://<myMaskingEngine>/masking/api-client/#!/tableMetadata/createTableMetadata>

Example

See how to use this in the context of a script [here](#).

Get All column metadata belonging to table metadata

Object references you will need:

- The table metadata ID to get the columns for

 This table metadata ID (2, in this example) is included in the QUERY STRING for this operation, NOT the payload.

REQUEST

```
curl -X GET --header 'Accept: application/json' --header 'Authorization:
7c856e3d-5b20-4261-b5fe-cc2ffcee5ae0'
'http://<myMaskingEngine>/masking/api/column-metadata?table_metadata_id=2'
```

RESPONSE

```
[ { "columnMetadataId": 12, "columnName": "schoolnme",
"tableMetadataId": 2, "columnLength": 50, "isMasked": false,
"isPrimaryKey": false, "isIndex": false, "isForeignKey": false }, ... ]
```

Note that the above response has been truncated due to its length for the purposes of this documentation.

More info

<http://<myMaskingEngine>/masking/api-client/#!/columnMetadata/getAllColumnMetadata>

Example

See how to use this in the context of a script [here](#).

Update column metadata with algorithm assignment

Object references you will need:

- Column metadata ID for the column you wish to update



Tip

This column metadata ID (20, in this example) is included in the PATH for this operation, NOT the payload.

- Since the names can vary in the API and UI, you should use the names obtained through the API (these may not align with the UI).
- Algorithm name
- Domain name

REQUEST

```
curl -X PUT --header 'Content-Type: application/json' --header 'Accept: application/json' --header 'Authorization: 7c856e3d-5b20-4261-b5fe-cc2ffcee5ae0' -d '{ "algorithmName": "AddrLine2Lookup", "domainName": "ADDRESS_LINE2", "isProfilerWritable": false }' 'http://<myMaskingEngine>/masking/api/column-metadata/20'
```

RESPONSE

```
{ "columnMetadataId": 20, "columnName": "l2_address", "tableMetadataId": 2, "algorithmName": "AddrLine2Lookup", "domainName": "ADDRESS_LINE2", "columnLength": 512, "isMasked": true, "isProfilerWritable": false, "isPrimaryKey": false, "isIndex": false, "isForeignKey": false, "domainAssignedBy": "admin_user" }
```

More info

<http://<myMaskingEngine>/masking/api-client/#!/columnMetadata/updateColumnMetadata>

Example

See how to use this in the context of a script [here](#).

API calls for creating and running masking jobs

Below are examples of requests you might enter and responses you might receive from the Masking API client. For commands specific to your masking engine, work with your interactive client at <http://<myMaskingEngine>/masking/api-client/>

 In all code examples, replace **<myMaskingEngine>** with the hostname or IP address of your virtual machine.

 HTTPS (SSL/TLS) is recommended, but for explanatory purposes these examples use insecure HTTP.

Creating a masking job

Object references you will need:

- The ID of the ruleset for which you wish to create the masking job

REQUEST

```
curl -X POST --header 'Content-Type: application/json' --header 'Accept: application/json' --header 'Authorization: e23bad24-8760-4091-a131-34f235d9b2d6' -d '{ "jobName": "some_masking_job", "rulesetId": 7, "jobDescription": "This example illustrates a MaskingJob with just a handful of the possible fields set. It is meant to exemplify a simple JSON body that can be passed to the endpoint to create a MaskingJob.", "feedbackSize": 100000, "onTheFlyMasking": false }' 'http://<myMaskingEngine>/masking/api/masking-jobs'
```

RESPONSE

```
{ "jobId": 1, "jobName": "some_masking_job", "rulesetId": 7, "createdBy": "Axistech", "createdTime": "2017-07-04T00:31:00.952+0000", "environmentId": 2, "feedbackSize": 100000, "jobDescription": "This example illustrates a MaskingJob with just a handful of the possible fields set. It is meant to exemplify a simple JSON body that can be passed to the endpoint to create a MaskingJob.", "maxMemory": 1024, "minMemory": 1024, "multiTenant": false, "numInputStreams": 1, "onTheFlyMasking": false }
```

 The response includes the ID of the newly created job (“jobId”).

More info

<http://<myMaskingEngine>/masking/api-client/#!/maskingJob/createMaskingJob>

Running a masking job

Create a new execution of a masking job.

Object references you will need:

- The ID of the job you want to run

REQUEST

```
curl -X POST --header 'Content-Type: application/json' --header 'Accept: application/json' --header 'Authorization: e23bad24-8760-4091-a131-34f235d9b2d6' -d '{"jobId": 1 }' 'http://<myMaskingEngine>/masking/api/executions'
```

RESPONSE

```
{ "executionId": 1, "jobId": 1, "status": "RUNNING" }
```

More info

<http://<myMaskingEngine>/masking/api-client/#!/execution/createExecution>

Checking the status of a masking job

Object references you will need:

- The ID of the execution you want to check (IN THE PATH)

 This execution id (1, in this example) is included in the PATH for this operation, NOT the payload.

REQUEST

```
curl -X GET --header 'Accept: application/json' --header 'Authorization: 8935f7f7-6de6-40ba-80d8-d8956b71248b' 'http://<myMaskingEngine>/masking/api/executions/1'
```

RESPONSE

```
{  "executionId": 1,  "jobId": 1,  "status": "SUCCEEDED",  "rowsMasked": 1000,  "rowsTotal": 1000,  "startTime": "2019-02-14T21:51:13.253+0000",
```

```
"endTime": "2019-02-14T21:51:54.956+0000"
}
```

More info

<http://<myMaskingEngine>/masking/api-client/#!/execution/getExecutionById>

Retrieving execution events related to a masking job

Object references you will need:

- The ID of the execution you want to check (as a URL parameter).

 This execution id (1, in this example) is specified as a URL parameter for this operation.

The execution-events endpoint returns execution events for a specified job execution. These execution events report failures or warnings associated with the masking job execution. NOT specifying the execution in the URL parameter will retrieve all execution events for all masking jobs.

REQUEST

```
curl -X GET --header 'Accept: application/json' --header 'Authorization:
8935f7f7-6de6-40ba-80d8-d8956b71248b'
'http://<myMaskingEngine>/masking/api/execution-events?execution_id=1&page_number=1'
```

RESPONSE

```
{
  "_pageInfo": {
    "numberOnPage": 1,
    "total": 1
  },
  "responseList": [
    {
      "executionEventId": 1,
      "executionId": 1,
      "eventType": "UNMASKED_DATA",
      "severity": "WARNING",
      "cause": "PATTERN_MATCH_FAILURE",
      "count": 1000,
      "timeStamp": "2019-02-14T21:51:51.790+0000",
      "executionComponentId": 1,
      "maskedObjectName": "RCHARS64_T1_0",
      "algorithmName": "DateShiftVariable"
    }
  ]
}
```

More info

<http://<myMaskingEngine>/masking/api-client/#!/execution-events/getAllExecutionEvents>

Retrieving non-conformant data samples associated with an execution Event

Object references you will need:

- The ID(s) of the execution event(s) you want to check (as one or more URL parameters).



This execution event id (1, in this example) is specified as a URL parameter for this operation.

The non-conformant-data-sample endpoint returns non-conformant data samples for a specified job execution event. These non-conformant data samples will report the data patterns that caused the non-conformant data execution event to help identify why data is not getting masked. NOT specifying an execution event in the URL parameter will retrieve all non-conformant data samples events for all masking jobs.

REQUEST

```
curl -X GET --header 'Accept: application/json' --header 'Authorization:
8935f7f7-6de6-40ba-80d8-d8956b71248b'
'http://<myMaskingEngine>/masking/api/non-conformant-data-sample?
execution_event_id=1&page_number=1'
```

RESPONSE

```
{
  "_pageInfo": {
    "numberOnPage": 7,
    "total": 7
  },
  "responseList": [
    {
      "dataSampleId": 1,
      "executionEventId": 1,
      "dataSample": "LLLLL",
      "count": 200
    },
    {
      "dataSampleId": 2,
      "executionEventId": 1,
      "dataSample": "LLLLLL",
      "count": 400
    },
    {
      "dataSampleId": 3,
      "executionEventId": 1,
      "dataSample": "LLLL",
      "count": 80
    }
  ]
}
```

```
},
{
  "dataSampleId": 4,
  "executionEventId": 1,
  "dataSample": "LLLLLLLL",
  "count": 100
},
{
  "dataSampleId": 5,
  "executionEventId": 1,
  "dataSample": "LLLLLLLLLLLLL",
  "count": 50
},
{
  "dataSampleId": 6,
  "executionEventId": 1,
  "dataSample": "LLLLLLLLLLLL",
  "count": 10
},
{
  "dataSampleId": 7,
  "executionEventId": 1,
  "dataSample": "LLLLLLLLLL",
  "count": 40
}
]
}
```

More info

<http://<myMaskingEngine>/masking/api-client/#!/non-conformant-data-sample/getAllNon-conformantDataSamples>

API calls involving file upload and download

File download

API calls involving file download through API client are noteworthy because if the request fails, the API client will continue to show the "loading" icon indefinitely.

To avoid this, make all file download calls through CURL instead. An example of a file download call using CURL is below.

```
curl -X GET --header 'Accept: application/octet-stream' --header
'Authorization: ec443730-124e-4958-a872-324a975bb500'
-o "/home/user/downloads"
'http://<myMaskingEngine>/masking/api/file-downloads/EXPORT-
ZXhwb3J0X2RvY3VtZW50X2dGZU9JMVYxLmpzb24%3D'
```

The `-o` flag from above specifies the location to save the file to.

File upload

API calls involving file upload are noteworthy because the generated curl from the Masking API client will be **missing the parameter referencing the file**; as such, those commands from the Masking API client **will not work**.

Instead, below are examples of working requests and responses for API calls involving file upload.

For commands specific to your masking engine, work with your interactive client at <http://<myMaskingEngine>/masking/api-client/>

 HTTPS (SSL/TLS) is recommended, but for explanatory purposes these examples use insecure HTTP.

 In all code examples, replace `<myMaskingEngine>` with the hostname or IP address of your virtual machine.

Creating a file format

REQUEST

```
curl -X POST --header 'Content-Type: multipart/form-data' --header
'Accept: application/json' --header 'Authorization:
d1313dd8-2ed9-4699-8e88-2b6a089ae2a6' -F
fileFormat=@/path/to/file_format/delimited_format.txt -F
fileFormatType=DELIMITED
'http://<myMaskingEngine>/masking/api/file-formats'
```

RESPONSE

```
{ "fileFormatId": 123, "fileFormatName": "delimited_format.txt",  
  "fileFormatType": "DELIMITED"  
}
```

More info

<http://<myMaskingEngine>/masking/api-client/#!/fileFormat/createFileFormat>

Creating an SSH Key

REQUEST

```
curl -X POST --header 'Content-Type: multipart/form-data' --header  
'Accept: application/json' --header 'Authorization:  
d1313dd8-2ed9-4699-8e88-2b6a089ae2a6' -F  
sshKey=@/path/to/ssh_key/this_file_name_is_your_ssh_key_name.txt  
'http://<myMaskingEngine>/masking/api/ssh-keys'
```

RESPONSE

```
{ "sshKeyName": "this_file_name_is_your_ssh_key_name.txt"  
}
```

More info

<http://<myMaskingEngine>/masking/api-client/#!/sshKey/createSshKey>

Backwards compatibility API usage

 In all examples, replace **<myMaskingEngine>** with the hostname or IP address of your virtual machine.

In all examples, replace **<myMaskingEngine>** with the hostname or IP address of your virtual machine.

API versioning context

The Masking API is versioned in accordance with the Semantic Versioning format: <http://semver.org/>. When the Masking API is updated, a new API version will be released. Scripts must reference an explicit API version or else there are no guarantees that the scripts will work on future releases of the Masking API.

Pinning down a version number to guarantee backwards-compatibility

'http://<myMaskingEngine>/masking/api/v5.0.0/environments'

This is the format for specifying a version in the URL of an API request targeting the **environments** endpoints.

Specifying the version for endpoint guarantees that the requester receives a response containing all of the fields that were present in that version of the API. This is intended to allow scripts that specify a masking API version in the URL to continue working upon future upgrades of the Masking product--even if a newer version of the API is available in the future Masking product.

For example, consider the scenario where a script is being developed today with a pinned down version **v5.0.0** in the URL of the API requests. Upon upgrade to a future release of the Masking product that has the API **v5.1.0** available, the same, untouched script that was developed with the pinned down version **v5.0.0** in the URL of the API requests are expected to continue working. That said, in order to leverage any new features of the API **v5.1.0**, the original script will need to be updated to specify the new API version in the URL, and the requests may need to be updated to conform to the new API specification.

While specifying a version for endpoint guarantees that all fields present in that version will be contained in the API response, it does **not** mean that new fields that have since been added to that endpoint in subsequent versions will be excluded. We, therefore, recommend that API users write their scripts to parse the JSON response objects by key name, rather than by key index, to prevent these additional fields from breaking any scripts.

Omitted version numbers

'http://<myMaskingEngine>/masking/api/environments'

This is the format for not specifying a version in the URL of an API request targeting the **environments** endpoints. When the API version number is omitted, the latest API version is taken as a default. In the first 5.2 release, an API request with an omitted version number will be interpreted as a request against the **v5.0.0** version of the API. In a future release that hypothetically has the API **v5.3.0** available, an API request with an omitted version number will be interpreted as a request against the **v5.3.0** version of the API.

Scripts that omit the version of the Masking API in the URL are not guaranteed to work upon future upgrades of the Masking product because the API specification may change between versions and requests that conform to the old API specification may not work on the new API specification.

DefaultApiVersion

If the version is omitted from the base path of the request's URL, but wishes to be treated using a specific masking API version that is not the latest version, set the DefaultApiVersion application setting. If the DefaultApiVersion is not set and the version is omitted from the URL, the latest version of the API on that engine will be used.

- The DefaultApiVersion application setting will not be applied to any requests made from within the masking engine. This mean that the UI, api-client, and phone home will always use the latest API version supported on the engine.

API response escaping

In Masking API responses, a backslash character (\) is escaped with an additional backslash character (\\).

Special attention should be paid to this behavior in scenarios where an API response is passed to another system as an input, for example, an automation system.

In such cases, a response might need special handling to convert the double backslash sequence (\\) back to a single backslash (\).

For example, consider the `POST /ssh-key` API for creating/installing an SSH Key. The result when the `POST /ssh-key` API is called with a file name that contains \ , such as `\key.txt` , is shown below.

Response Body:

```
{
  "errorMessage": "SSH Key file name should not contain [\\, ;, %, ?, :]"
}
```

API call for generating support bundle

Below are examples of requests you might enter and responses you might receive from the Masking API client. For commands specific to your masking engine, work with your interactive client at <http://<myMaskingEngine>/masking/api-client/>

 In all code examples, replace **<myMaskingEngine>** with the hostname or IP address of your virtual machine.

 HTTPS (SSL/TLS) is recommended, but for explanatory purposes these examples use insecure HTTP.

Generating a support bundle

No arguments are required for launching a support bundle generation task

REQUEST

```
curl -X POST --header 'Content-Type: application/json' --header 'Accept: application/json' --header 'Authorization: 5f745517-d4ce-45de-afb3-6be06205188f' 'http://<myMaskingEngine>/masking/api/v5.2.0/support-bundle'
```

RESPONSE

```
{
  "asyncTaskId": 5,
  "operation": "SUPPORT_BUNDLE_GENERATE",
  "reference": "",
  "status": "RUNNING",
  "startTime": "2022-06-02T16:33:42.792+00:00",
  "cancellable": true
}
```

 The response includes the ID of the launched asynchronous task (“asyncTaskId”) which runs the scripts collecting the Support Bundle information and packs it into tar.gz file.

Reading the async task result

Object references you will need:

- The ID of the asyncTask retrieved from the response of the `supportBundle` endpoint (in the response example above equals 5)

Retrieve the result of the async task by running `asyncTask GET /async-tasks/{asyncTaskId}` endpoint.

REQUEST

```
curl -X GET --header 'Accept: application/json' --header
'Authorization: 5f745517-d4ce-45de-afb3-6be06205188f'
'http://<myMaskingEngine>/masking/api/v5.2.0/async-tasks/6'
```

- ⊞ Gathering the support bundle information might be a relatively long task, running minutes (depending on the amount of the accumulated information on the Masking Engine). While it's not finished the response will have "status": "RUNNING" and "reference":"" (i.e. empty).

After the task is finished the response will look like:

```
{
  "asyncTaskId": 6,
  "operation": "SUPPORT_BUNDLE_GENERATE",
  "reference": "SUPPORT_BUNDLE-dlpx-support-564db0c0-162b-c22f-f2ed-
b17ff6b933b0-20220602-16-48-03.tar.gz",
  "status": "SUCCEEDED",
  "startTime": "2022-06-02T16:48:02.969+00:00",
  "endTime": "2022-06-02T16:50:25.960+00:00",
  "cancellable": true
}
```

Retrieving the generated support bundle file

The Support Bundle file may be downloaded using the `fileDownload GET /file-downloads/{fileDownloadId}` API endpoint.

Object references you will need:

- The reference provided in the succeeded async task response, to be used as `fileDownloadId` input argument.

- ⊞ Getting the Support Bundle file to the browser memory might take few minutes (depending on the generated Support Bundle size).

- ⊞ The `Response Content Type` field should be set to `application/octet-stream` value. If it's left on the default `application/json` than the downloaded file wouldn't be recognized as a valid tar file.

REQUEST

```
curl -X GET --header 'Accept: application/octet-stream'
--header 'Authorization: 5f745517-d4ce-45de-afb3-6be06205188f'
```

```
'http://<myMaskingEngine>/masking/api/v5.2.0/file-downloads/SUPPORT_BUNDLE-dlpx-support-564db0c0-162b-c22f-f2ed-b17ff6b933b0-20220602-16-48-03.tar.gz'
```

RESPONSE

The `Response Body` is represented as a clickable download URL, for example: `Download`

```
SUPPORT_BUNDLE-dlpx-support-564db0c0-162b-c22f-f2ed-b17ff6b933b0-20220602-16-48-03.tar.gz
```

Here you have 2 options:

1. Click on that link, and the Support Bundle tar file would be downloaded to your default download directory.
2. Use the above curl command to download the support bundle file. To keep the same file name you need to add `-O` (capital letter O) argument to this curl command, for example:

```
$ curl -X GET --header 'Accept: application/octet-stream' --header 'Authorization: 5f745517-d4ce-45de-afb3-6be06205188f' 'http://<myMaskingEngine>/masking/api/v5.2.0/file-downloads/SUPPORT_BUNDLE-dlpx-support-564db0c0-162b-c22f-f2ed-b17ff6b933b0-20220602-16-48-03.tar.gz' -O
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left   Speed
 100 17.5M    0 17.5M    0     0    735k      0  --:--:--  0:00:24  --:--:--  782k
```

or

```
$ curl -X GET --header 'Accept: application/octet-stream' --header 'Authorization: 5f745517-d4ce-45de-afb3-6be06205188f' 'http://<myMaskingEngine>/masking/api/v5.2.0/file-downloads/SUPPORT_BUNDLE-dlpx-support-564db0c0-162b-c22f-f2ed-b17ff6b933b0-20220602-16-48-03.tar.gz' --output support_bundle.tar.gz
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left   Speed
 100 17.5M    0 17.5M    0     0    660k      0  --:--:--  0:00:27  --:--:--  426k
```

I If you choose not to use the curl command but clicking to the download URL - the downloaded file has autogenerated suffix added to the name of the file, for example: `application_octet-stream_SUPPORT_BUNDLE-dlpx-support-564db0c0-162b-c22f-f2ed-b17ff6b933b0-20220602-16-48-03.tar.gz_blob_http___<>` You might use that file as is, or rename it to the desired name. The recommendation is to leave the `.tar.gz` extension.

More info

- Only one support bundle generation task can be running at a time.
- Support Bundle generation is cancellable (via `asyncTask PUT /async-tasks/{asyncTaskId}/cancel` endpoint).

API examples

This section covers the following topics:

- [loginCredentials](#)
- [helpers](#)
- [apiHostInfo](#)
- [Configure enclosure escape character](#)
- [createApplication](#)
- [createEnvironment](#)
- [createInventory](#)
- [create DatabaseConnector](#)
- [create DatabaseRuleset](#)
- [getBillingUsage](#)
- [getAuditLogs](#)
- [getSyncableObjects](#)
- [getSyncableObjectsExport](#)
- [profileTypeExpressions](#)
- [runMaskingJob](#)
- [getDatabaseUsage](#)

loginCredentials

```
#!/bin/bash

#
# This file contains all the login information for the masking engine.
#

# Login credentials for the Masking Engine.
USERNAME="myUsername"
PASSWORD="myPassword"

# Login into a masking engine
curl -X POST --header 'Content-Type: application/json' --header 'Accept: application/
json' -d '{ \ "username": "myUsername", \ "password": "myPassword" \ }' 'http://
<myMaskingEngine>/masking/api/login'
```

helpers

```
#!/bin/bash

#
# This file contains helpers for the various Masking API cookbook scripts.
# This script uses jq to process JSON. More information can be found here - https://stedolan.github.io/jq/.
#

# Login and set the correct $AUTH_HEADER.
login() {
    echo "* logging in..."
    LOGIN_RESPONSE=$(curl -s $SSL_CERT -X POST -H 'Content-Type: application/json' -H
'Accept: application/json' --data @- $MASKING_ENGINE/login <<EOF
{
    "username": "$USERNAME",
    "password": "$PASSWORD"
}
EOF
) || die "Login failed with exit code $?"
    check_error "$LOGIN_RESPONSE"
    TOKEN=$(echo $LOGIN_RESPONSE | jq -r '.Authorization')
    AUTH_HEADER="Authorization: $TOKEN"
}

# Get all applications and select the first one. Place the applicationName in
$APPLICATION_ID.
get_application_id() {
    echo "* getting all applications and selecting first one"
    APPLICATIONS_RESPONSE=$(curl -s $SSL_CERT -X GET -H "'$AUTH_HEADER'" -H
'Content-Type: application/json' $MASKING_ENGINE/applications)
    check_error "$APPLICATIONS_RESPONSE"
    NUM_APPLICATIONS=$(echo $APPLICATIONS_RESPONSE | jq -r '._pageInfo.total')
    check_empty $NUM_APPLICATIONS "found no applications to use"
    APPLICATION_ID=$(echo $APPLICATIONS_RESPONSE | jq -r
'.responseList[0].applicationName')
    echo "using application '$APPLICATION_ID'"
}

# Get all environments and select the first one. Place the environmentId in
$ENVIRONMENT_ID.
get_environment_id() {
    echo "* getting all environments and selecting first one"
    ENVIRONMENTS_RESPONSE=$(curl -s $SSL_CERT -X GET -H "'$AUTH_HEADER'" -H
'Content-Type: application/json' $MASKING_ENGINE/environments)
    check_error "$ENVIRONMENTS_RESPONSE"
    NUM_ENVIRONMENTS=$(echo $ENVIRONMENTS_RESPONSE | jq -r '._pageInfo.total')
    check_empty $NUM_ENVIRONMENTS "found no environments to use"
    ENVIRONMENT_ID=$(echo $ENVIRONMENTS_RESPONSE | jq -r
'.responseList[0].environmentId')
```

```

    echo "using environment '$ENVIRONMENT_ID'"
}
# Get all database connectors and select the first one. Place the databaseConnectorId
in $CONNECTOR_ID.
get_connector_id() {
    echo "* getting all database connectors and selecting first one"
    CONNECTORS_RESPONSE=$(curl -s $SSL_CERT -X GET -H ""$AUTH_HEADER"" -H 'Content-
Type: application/json' $MASKING_ENGINE/database-connectors)
    check_error "$CONNECTORS_RESPONSE"
    NUM_CONNECTORS=$(echo $CONNECTORS_RESPONSE | jq -r '._pageInfo.total')
    check_empty $NUM_CONNECTORS "found no db connectors to use"
    CONNECTOR_ID=$(echo $CONNECTORS_RESPONSE | jq -r
'.responseList[0].databaseConnectorId')
    echo "using database connector '$CONNECTOR_ID'"
}

# Get all database rulesets and select the first one. Place the databaseRulesetId in
$RULESET_ID.
get_ruleset_id() {
    echo "* getting all database rulesets and selecting first one"
    RULESETS_RESPONSE=$(curl -s $SSL_CERT -X GET -H ""$AUTH_HEADER"" -H 'Content-
Type: application/json' $MASKING_ENGINE/database-rulesets)
    check_error "$RULESETS_RESPONSE"
    NUM_RULESETS=$(echo $RULESETS_RESPONSE | jq -r '._pageInfo.total')
    check_empty $NUM_RULESETS "found no db rulesets to use"
    RULESET_ID=$(echo $RULESETS_RESPONSE | jq -r '.responseList[0].databaseRulesetId')
    echo "using database ruleset '$RULESET_ID'"
}

# Get all database tables for a database connector specified by $CONNECTOR_ID. Select
the first one and place in $TABLE_NAME.
get_table() {
    echo "* getting all tables for connector '$CONNECTOR_ID' and selecting first one"
    TABLES_RESPONSE=$(curl -s $SSL_CERT -X GET -H ""$AUTH_HEADER"" -H 'Content-
Type: application/json' $MASKING_ENGINE/database-connectors/$CONNECTOR_ID/fetch)
    check_error "$TABLES_RESPONSE"
    NUM_TABLES=$(echo $TABLES_RESPONSE | jq -r '. | length')
    check_empty $NUM_TABLES "found no tables to use"
    TABLE_NAME=$(echo $TABLES_RESPONSE | jq -r '.[0]')
    echo "using table '$TABLE_NAME'"
}

# Get all column metadata for table metadata specified by $TABLE_METADATA_ID. Select
the first one and place in $COLUMN_METADATA_ID.
get_column_metadata_id() {
    echo "* getting all column metadata belonging to table metadata
'$TABLE_METADATA_ID' and selecting the first one"
    COLUMNS_RESPONSE=$(curl -s $SSL_CERT -X GET -H ""$AUTH_HEADER"" -H 'Content-
Type: application/json' $MASKING_ENGINE/column-metadata?)
    table_metadata_id=$TABLE_METADATA_ID
    check_error "$COLUMNS_RESPONSE"
    NUM_COLUMNS=$(echo $COLUMNS_RESPONSE | jq -r '. | length')
    check_empty $NUM_COLUMNS "found no columns to use"
}

```

```

COLUMN_METADATA=$(echo $COLUMNS_RESPONSE | jq -r '.responseList[0]')
COLUMN_METADATA_ID=$(echo $COLUMN_METADATA | jq -r '.columnMetadataId')
echo "using column '$COLUMN_METADATA_ID'"
}

# Get all masking jobs and select the first one. Place the jobId in $MASKING_JOB_ID.
get_masking_job_id() {
  echo "* getting all masking jobs and selecting first one"
  MASKINGJOB_RESPONSE=$(curl -s $SSL_CERT -X GET -H ""$AUTH_HEADER"" -H 'Content-
Type: application/json' $MASKING_ENGINE/masking-jobs)
  check_error "$MASKINGJOB_RESPONSE"
  NUM_MASKINGJOB=$(echo $MASKINGJOB_RESPONSE | jq -r '._pageInfo.total')
  check_empty $NUM_MASKINGJOB "found no masking jobs to use"
  MASKING_JOB_ID=$(echo $MASKINGJOB_RESPONSE | jq -r
'.responseList[0].maskingJobId')
  echo "using masking job '$MASKINGJOB_ID'"
}

# run_masking_job and save execution id in $MASKING_EXECUTION_ID.
run_masking_job() {
  echo "* running masking job '$MASKING_JOB_ID'..."
  MASKINGJOB_RESPONSE1=$(curl $SSL_CERT -X POST -H ""$AUTH_HEADER"" -H 'Content-
Type: application/json' -H 'Accept: application/json' --data @- $MASKING_ENGINE/
executions <<EOF
  {
    "jobId": "$MASKING_JOB_ID"
  }
EOF
)
  echo "Response for Masking job is: '$MASKINGJOB_RESPONSE1'"
  MASKING_EXECUTION_ID=$(echo $MASKINGJOB_RESPONSE1 | jq -r '.executionId')
}

# get_execution_status in $MASKING_EXECUTION_STATUS.
get_execution_status() {
  echo "* Getting execution details.....for execution id = $1"
  MASKINGJOB_RESPONSE=$(curl -s $SSL_CERT -X GET -H ""$AUTH_HEADER"" -H 'Content-
Type: application/json' $MASKING_ENGINE/executions/$1)
  check_error "$MASKINGJOB_RESPONSE"
  MASKING_EXECUTION_STATUS=$(echo $MASKINGJOB_RESPONSE | jq -r '.status')
  echo "Execution status for id= $1 is '$MASKING_EXECUTION_STATUS'"
}

# Check if $1 is equal to 0. If so print out message specified in $2 and exit.
check_empty() {
  if [ $1 -eq 0 ]; then
    echo $2
    exit 1
  fi
}

```

```
# Check if $1 is an object and if it has an 'errorMessage' specified. If so, print
the object and exit.
check_error() {
    # jq returns a literal null so we have to check against that...
    if [ "$(echo "$1" | jq -r 'if type=="object" then .errorMessage else "null"
end')" != 'null' ]; then
        echo $1
        exit 1
    fi
}

# Print the message and exit the program.
die() {
    echo
    "*****"
    echo "$(basename $0) ERROR: $*" >&2
    echo
    "*****"
    exit 1
}
```

apiHostInfo

```
#!/bin/bash

#
# This file contains all the host information for the masking engine. Additionally,
# this file allows configuration of SSL if desired.
#

# update host name
HOST="myMaskingEngine.com"
API_PATH="masking/api"

# To connect via SSL, set $SSL to "on" and update the port if necessary (default 8443)
.
# Additionally, you must update the path to the ssl certificate.
SSL="off"
SSL_PORT="8443"
# update cert name
SSL_CERT_PATH="self-signed.cer"

if [ "$SSL" = "on" ]
then
    MASKING_ENGINE="https://$HOST:$SSL_PORT/$API_PATH"
    SSL_CERT="--cacert $SSL_CERT_PATH"
else
    MASKING_ENGINE="http://$HOST/$API_PATH"
    SSL_CERT=""
fi
```

Configure enclosure escape character

```
#!/bin/bash

#
# This script uses the Masking Engine APIs to configure the enclosure escape
# character feature.
# The script uses the /login API to obtain an authentication token and then uses the
# PUT /file-metadata API.
#
# To use this script, you must set DOUBLE_ENCLOSURE,
# CUSTOM_ENCLOSURE_ESCAPE_CHARACTER, ESCAPE_ENCLOSURE_ESCAPE_CHARACTER and RULESET_ID
# accordingly
#

source ./apiHostInfo.bash
eval $(cat ./loginCredentials.bash)
source ./helpers.bash

helpFunction() {
    echo ""
    echo "Usage: $0 -h HELP -d DOUBLE_ENCLOSURE -c CUSTOM_ENCLOSURE_ESCAPE_CHARACTER
-e ESCAPE_ENCLOSURE_ESCAPE_CHARACTER -r RULESET_ID"
    echo -e "\t-h Show the usage of the script"
    echo -e "\t-d Set the value for DOUBLE_ENCLOSURE"
    echo -e "\t-c Set the value for CUSTOM_ENCLOSURE_ESCAPE_CHARACTER"
    echo -e "\t-e Set the value for ESCAPE_ENCLOSURE_ESCAPE_CHARACTER"
    echo -e "\t-r Set the value for RULESET_ID"
    echo -e "\n\tAdditional Note:"
    echo -e "\t1: The default value for parameter D=true, no need to set the value if
you want to set enclosure escape character same as enclosure character."
    echo -e "\t2: If parameter D=true then custom enclosure escape character value
will be ignored."
    echo -e "\t3: The default value for parameter E=false, change accordingly as per
the requirement."
    echo -e "\t4: If parameter R is blank, it means changes will be applicable for
all rulesets. Pass the R={RULESET_ID} if you want to update the settings only for the
given ruleset. Example R=1"
    exit 1 # Exit script after printing help
}

# Set DOUBLE_ENCLOSURE=true if you want to set enclosure escape character same as
# enclosure character,
# and if DOUBLE_ENCLOSURE=true then CUSTOM_ENCLOSURE_ESCAPE_CHARACTER value will be
# ignored.
DOUBLE_ENCLOSURE=true
# Replace * with your custom escape character if you want to set custom enclosure
# escape character
# and also DOUBLE_ENCLOSURE=false need to set
CUSTOM_ENCLOSURE_ESCAPE_CHARACTER="*"

```

```

# Modify ESCAPE_ENCLOSURE_ESCAPE_CHARACTER value accordingly.
ESCAPE_ENCLOSURE_ESCAPE_CHARACTER=false
# Comment this RULESET_ID if you want to update for all delimited file ruleset for
  which enclosure is defined.
#RULESET_ID=1

while getopts "h:d:c:e:r:" opt; do
  case "$opt" in
    h) helpFunction exit ;;
    d) DOUBLE_ENCLOSURE="$OPTARG" ;;
    c) CUSTOM_ENCLOSURE_ESCAPE_CHARACTER="$OPTARG" ;;
    e) ESCAPE_ENCLOSURE_ESCAPE_CHARACTER="$OPTARG" ;;
    r) RULESET_ID="$OPTARG" ;;
    ?) helpFunction ;; # Print helpFunction in case parameter is non-existent
  esac
done

# Print helpFunction in case parameters are empty
if [ -z "$DOUBLE_ENCLOSURE" ] || [ -z "$CUSTOM_ENCLOSURE_ESCAPE_CHARACTER" ] || [ -z
"$ESCAPE_ENCLOSURE_ESCAPE_CHARACTER" ]; then
  echo "Some or all of the parameters are empty"
  helpFunction
fi

login

echo "Calling GET /file-metadata API"
if [[ -z "$RULESET_ID" ]] || [ "$RULESET_ID" = "null" ] || [ "$RULESET_ID" = "" ];
then
  echo "Configuring the enclosure escape character feature for all File Ruleset."
  FILE_METADATA_RESPONSE=$(curl $SSL_CERT -X GET -H ""$AUTH_HEADER"" -H 'Accept:
application/json' ""$MASKING_ENGINE/file-metadata")
else
  echo "Configuring the enclosure escape character feature for File
Ruleset(RULESET_ID=$RULESET_ID)"
  FILE_METADATA_RESPONSE=$(curl $SSL_CERT -X GET -H ""$AUTH_HEADER"" -H 'Accept:
application/json' ""$MASKING_ENGINE/file-metadata?ruleset_id=$RULESET_ID")
fi

i=0
while true; do
  ENCLOSURE=$(jq '.responseList[$i] .enclosure' <<<"$FILE_METADATA_RESPONSE")

  if [ "$DOUBLE_ENCLOSURE" = true ]; then
    CUSTOM_ENCLOSURE_ESCAPE_CHARACTER=$ENCLOSURE
  fi

  UPDATED_FILE_METADATA_RESPONSE=$(jq '.responseList[$i] .enclosureEscapeCharacte
r='$CUSTOM_ENCLOSURE_ESCAPE_CHARACTER' <<<"$FILE_METADATA_RESPONSE")
  UPDATED_FILE_METADATA_RESPONSE=$(jq '.responseList[$i] .escapeEnclosureEscapeCh
aracter='$ESCAPE_ENCLOSURE_ESCAPE_CHARACTER' <<<"$UPDATED_FILE_METADATA_RESPONSE")
  FILE_METADATA_RESPONSE=$UPDATED_FILE_METADATA_RESPONSE

```

```

FILE_METADATA_OBJECT=$(jq '.responseList['$i']' <<<"$FILE_METADATA_RESPONSE")
FILE_METADATA_ID=$(jq '.responseList['$i'] .fileMetadataId' <<<"$FILE_METADATA_RE
SPONSE")

if [[ -z "$FILE_METADATA_ID" ]] || [ "$FILE_METADATA_ID" = "null" ]; then
    break
else
    if [[ ! -z "$ENCLOSURE" ]] && [ ! "$ENCLOSURE" = "null" ] && [ ! "$ENCLOSURE"
= "" ]; then
        echo "Calling $MASKING_ENGINE/file-metadata/$FILE_METADATA_ID API to
update enclosureEscapeCharacter=$CUSTOM_ENCLOSURE_ESCAPE_CHARACTER and
escapeEnclosureEscapeCharacter=$ESCAPE_ENCLOSURE_ESCAPE_CHARACTER"
        UPDATE_RESPONSE=$(curl $SSL_CERT -X PUT -H '"$AUTH_HEADER"' -H
'Content-Type: application/json' -H 'Accept: application/json' -d '"$FILE_METADATA_O
BJECT"' '"$MASKING_ENGINE/file-metadata/$FILE_METADATA_ID"')
        check_error "$UPDATE_RESPONSE"
    fi
fi
((i++))
done
echo

```

createApplication

```
#!/bin/bash

#
# This script will login and create an application. It depends on helpers in the
# helpers script as well as host and login
# information found in apiHostInfo and loginCredentials, respectively.
#

source apiHostInfo
eval $(cat loginCredentials)
source helpers

login

echo "* creating application 'App123'..."
curl $SSL_CERT -X POST -H ""$AUTH_HEADER"" -H 'Content-Type: application/json' -H
'Accept: application/json' --data @- $MASKING_ENGINE/applications <<EOF
{
    "applicationName": "App123"
}
EOF

echo
```

createEnvironment

```
#!/bin/bash

#
# This script will login and create an environment with an application. It depends on
# helpers in the helpers
# script as well as host and login information found in apiHostInfo and
# loginCredentials, respectively.
#

source apiHostInfo
eval $(cat loginCredentials)
source helpers

login

#
# When deciding which application to place the environment in we simply choose the
# first application found. You are
# encouraged to modify this to suit your needs. Please see get_application_id in
# helpers for more information.
#
get_application_id

echo "* creating environment 'newEnv' in application '$APPLICATION_ID'..."
curl $SSL_CERT -X POST -H '$AUTH_HEADER' -H 'Content-Type: application/json' -H
'Accept: application/json' --data @- $MASKING_ENGINE/environments <<EOF
{
  "environmentName": "newEnv",
  "application": "$APPLICATION_ID",
  "purpose": "MASK"
}
EOF

echo
```

createInventory

```
#!/bin/bash

#
# This script will login and create an environment with an application. It depends on
helpers in the helpers
# script as well as host and login information found in apiHostInfo and
loginCredentials, respectively.
#

source apiHostInfo
eval $(cat loginCredentials)
source helpers

login

#
# When deciding which application to place the environment in we simply choose the
first application found. You are
# encouraged to modify this to suit your needs. Please see get_application_id in
helpers for more information.
#
get_application_id

echo "* creating environment 'newEnv' in application '$APPLICATION_ID'..."
curl $SSL_CERT -X POST -H '$AUTH_HEADER' -H 'Content-Type: application/json' -H
'Accept: application/json' --data @- $MASKING_ENGINE/environments <<EOF
{
    "environmentName": "newEnv",
    "application": "$APPLICATION_ID",
    "purpose": "MASK"
}
EOF

echo
```

create DatabaseConnector

```
#!/bin/bash

#
# This script will login and create a database connector in an environment. It
# depends on helpers in the helpers
# script as well as host and login information found in apiHostInfo and
# loginCredentials, respectively.
#

source apiHostInfo
eval $(cat loginCredentials)
source helpers

login

#
# When deciding which environment to place the connector in we simply choose the
# first environment found. You are
# encouraged to modify this to suit your needs. Please see get_environment_id in
# helpers for more information.
#
get_environment_id

echo "* creating database connector 'connector' in environment '$ENVIRONMENT_ID'..."
curl $SSL_CERT -X POST -H '$AUTH_HEADER' -H 'Content-Type: application/json' -H
'Accept: application/json' --data @- $MASKING_ENGINE/database-connectors <<EOF
{
    "connectorName": "connector",
    "databaseType": "ORACLE",
    "environmentId": $ENVIRONMENT_ID,
    "host": "myHost",
    "password": "myPassword",
    "port": 1234,
    "schemaName": "MYSHEMA",
    "sid": "mySID",
    "username": "MYUSERNAME"
}
EOF

echo
```

create DatabaseRuleset

```
#!/bin/bash

#
# This script will login and create a database ruleset for a database connector. It
# depends on helpers in the helpers
# script as well as host and login information found in apiHostInfo and
# loginCredentials, respectively.
#

source apiHostInfo
eval $(cat loginCredentials)
source helpers

login

#
# When deciding which database connector we will use, we simply choose the first
# database connector found. You are
# encouraged to modify this to suit your needs. Please see get_connector_id in
# helpers for more information.
#
get_connector_id

echo "* creating database ruleset 'myRuleset' in db connector '$CONNECTOR_ID'..."
curl $SSL_CERT -X POST -H '$AUTH_HEADER' -H 'Content-Type: application/json' -H
'Accept: application/json' --data @- $MASKING_ENGINE/database-rulesets <<EOF
{
  "rulesetName": "myRuleset",
  "databaseConnectorId": $CONNECTOR_ID
}
EOF

echo
```

getBillingUsage

```
#!/bin/bash

#
# This script is an "out of the box" script that goes through
# Login and GET /billing-usage with the authentication
# token from Login
#

source apiHostInfo
eval $(cat loginCredentials)
source helpers

login

START_DATE=yyyy-mm-dd
END_DATE=yyyy-mm-dd

echo "* GET /billing-usage for date range $START_DATE - $END_DATE from
$EXPORT_ENGINE"
EXPORT_RESPONSE=$(curl $SSL_CERT -X GET -H '"$AUTH_HEADER"' -H 'Accept:
application/json' $MASKING_ENGINE/billing-usage?
start_date=$START_DATE&end_date=$END_DATE) || die "curl failed with exit code $?"
echo $EXPORT_RESPONSE
```

getAuditLogs

```
#!/bin/bash

#
# This script is an "out of the box" script that goes through
# Login and GET /audit-logs with the authentication
# token from Login
#

source apiHostInfo
eval $(cat loginCredentials)
source helpers

login

echo "* GET /audit-logs from $EXPORT_ENGINE"
EXPORT_RESPONSE=$(curl $SSL_CERT -X GET -H '"$AUTH_HEADER"' -H 'Accept:
application/json' $MASKING_ENGINE/audit-logs)

# Calculate the number of audit log entries and the proximity to the entry limit.
AUDIT_ENTRY_COUNT=$(jq '._pageInfo.total' <<<"$EXPORT_RESPONSE")
MAX_ENTRIES=1000000
DIFFERENCE=$((MAX_ENTRIES-AUDIT_ENTRY_COUNT))

# Retrieve the date of the oldest audit entry retained.
OLDEST_DATE=$(jq '.responseList[1].activityTime' <<<"$EXPORT_RESPONSE")

echo "There are $AUDIT_ENTRY_COUNT entries in the audit log. After $DIFFERENCE more
audits you will hit the $MAX_ENTRIES limit and will begin to overwrite entries
starting from the oldest, which was created on: $OLDEST_DATE"
```

getSyncableObjects

```
#!/bin/bash

#
# This script is an "out of the box" script that goes through
# Login and GET /syncable-objects with the authentication
# token from Login
#

source apiHostInfo
eval $(cat loginCredentials)
source helpers

login

echo "* GET /syncable-objects from $EXPORT_ENGINE"
EXPORT_RESPONSE=$(curl $SSL_CERT -X GET -H '"$AUTH_HEADER"' -H 'Accept:
application/json' $MASKING_ENGINE/syncable-objects)
echo $EXPORT_RESPONSE
```

getSyncableObjectsExport

```
#!/bin/bash

#
# This script will log in and get all syncable objects on
# the Masking Engine and then, given a grouping command, save the
# exported document in a file and export all syncable objects
# in the indicated group
#
# Grouping command:
# algoType: -t <LOOKUP | BINARYLOOKUP | SEGMENT | TOKENIZATION | KEY>
# algoCd: -n <RegexForAlgoName>
#
# Currently the response from GET /syncable-objects is saved
# to getobj_response.json, and the grouped input for /export
# in grouped_export_list.json, and the final export response
# into export_response.json. But of course, this can script
# can be modified to save to other specified places.
#

source apiHostInfo
eval $(cat loginCredentials)
source helpers

login

echo "* GET /syncable-objects"
GETOBJ_RESPONSE=$(curl $SSL_CERT -X GET -H ""$AUTH_HEADER"" -H 'Content-Type:
application/json' $MASKING_ENGINE/syncable-objects)
echo $GETOBJ_RESPONSE > "./getobj_response.json"

# Create a temporary export list file
GROUPED_EXPORT_LIST="./grouped_export_list.json"
echo "[]" > $GROUPED_EXPORT_LIST

if [[ $1 == "-t" ]]; then
    ALGO_TYPE=$2
    echo "* Filter for all syncable objects of algorithm type $ALGO_TYPE"

    jq -c '.responseList[]' getobj_response.json | while read i; do
        if [[ $(echo $i | jq '.objectType') == \"$ALGO_TYPE\" ]]; then
            # The key to getting the correct json format here was to use
            # the --argjson instead of --arg. --arg will stringify everything
            # and escape all special characters like {, ", etc.
            echo $(cat $GROUPED_EXPORT_LIST | jq --argjson obj "$i" '. |= . + [$obj]') >
$GROUPED_EXPORT_LIST
        fi
    done
elif [[ $1 == "-n" ]]; then
```

```
ALGO_NAME_REGEX=$2
echo "* Filter for all syncable objects where algorithmCd matches the regex
$ALGO_NAME_REGEX"

jq -c '.responseList[]' getobj_response.json | while read i; do
    if [[ "$(echo $i | jq '.objectIdentifier.algorithmName')" =~
"$ALGO_NAME_REGEX" ]]; then
        echo $(cat $GROUPED_EXPORT_LIST | jq --argjson obj "$i" '. |= . + [$obj]')
> $GROUPED_EXPORT_LIST
    fi
done
fi

echo "* Export syncable objects from $GROUPED_EXPORT_LIST"
EXPORT_RESPONSE=$(curl $SSL_CERT -X POST -H ""$AUTH_HEADER"" -H 'Content-Type:
application/json' -H 'Accept: application/json' -d "<$GROUPED_EXPORT_LIST)"
$MASKING_ENGINE/export)

# Save the grouped export response into a file
echo $EXPORT_RESPONSE > export_response.json
echo '* Completed exporting. Check "export_response.json" for the export document.
This export document json object will be what you literally put in as the input for
import'
```

profileTypeExpressions

Add a new type expression

```
#!/bin/bash

#
# This script will login and create a profile type expression. It depends on helpers
# in the helpers script as well as host and login
# information found in apiHostInfo and loginCredentials, respectively.
#

source apiHostInfo
eval $(cat loginCredentials)
source helpers

login

curl $SSL_CERT -X POST -H ""$AUTH_HEADER"" -H 'Content-Type: application/json' -H
'Accept: application/json' --data @- $MASKING_ENGINE/profile-type-expressions <<EOF
{
  "domainName": "FIRST_NAME",
  "expressionName": "FirstNameType",
  "dataType": "String",
  "minDataLength": 5
}
EOF

echo
```

To be effective, a Profile Type Expression has to be part of a profile set. A type expression can be added to a profile set with the profile-sets endpoint. For example, if some Profile Type Expressions were created and have ids 57 and 48, we can use the PUT method on the profile-set endpoint to update an existing profile set so that it includes the new profile type expression. This is shown below, where the profile set has id 42.

```
#!/bin/bash

source apiHostInfo
eval $(cat loginCredentials)
source helpers

login

curl $SSL_CERT -X PUT -H ""$AUTH_HEADER"" -H 'Content-Type: application/json' -H
'Accept: application/json' --data @- $MASKING_ENGINE/profile-sets/42 <<EOF
{
  "profileSetName": "FINDS_ALL_SENSITIVE_DATA",
  "profileExpressionIds": [
    4,
```

```
    8,  
    12,  
    13,  
    27  
  ],  
  "profileTypeExpressionIds": [  
    57,  
    58  
  ]  
}  
EOF
```

Delete a type expression

Deleting a type expression is done using the DELETE method on the profile-type-expression endpoint. The expression must be removed from any profile sets it's a part of before it can be deleted.

```
#!/bin/bash  
  
#  
# This script will login and delete a profile type expression. It depends on helpers  
# in the helpers script as well as host and login  
# information found in apiHostInfo and loginCredentials, respectively.  
#  
  
source apiHostInfo  
eval $(cat loginCredentials)  
source helpers  
  
login  
  
echo "* creating application 'App123'..."  
curl $SSL_CERT -X DELETE -H ""$AUTH_HEADER"" -H 'Content-Type: application/json' -H  
'Accept: application/json' --data @- $MASKING_ENGINE/profile-type-expressions/57  
  
echo
```

runMaskingJob

This script will login and run a masking job. It depends on helpers in the helpers script as well as host and login information found in apiHostInfo and loginCredentials, respectively.

```
#!/bin/bash
source apiHostInfo
eval $(cat loginCredentials)
source helpers

login
```

When deciding which masking job to run, we simply choose the first masking job found. You are encouraged to modify this to suit your needs. Please see `get_masking_job_id` in helpers for more information.

```
get_masking_job_id

echo "* running masking job '$MASKING_JOB_ID'..."
curl $SSL_CERT -X POST -H ""$AUTH_HEADER"" -H 'Content-Type: application/json' -H
'Accept: application/json' --data @- $MASKING_ENGINE/executions <<EOF
{
  "jobId": "$MASKING_JOB_ID"
}
EOF
echo
```

If a masking job is called by a PowerShell hook script, the following command **MUST** be added to the script using the **PowerShell -File** prefix, **file path**, and the **exit \$LASTEXITCODE** suffix.

```
PowerShell -File C:\Users\HomeFolder\AddUser.ps1; exit $LASTEXITCODE
```

If this is not added then Delphix will not know if the script ran or completed. For more information, please visit this [SQL Server PowerShell Script Error Handling](#) documentation.

getDatabaseUsage

```
#!/bin/bash

#
# This script is an "out of the box" script that goes through
# Login and GET /database-usage with the authentication
# token from Login
#

source apiHostInfo
eval $(cat loginCredentials)
source helpers

login

echo "* GET /database-usage from $EXPORT_ENGINE"
EXPORT_RESPONSE=$(curl $SSL_CERT -X GET -H "$AUTH_HEADER" -H 'Accept:
application/json' $MASKING_ENGINE/database-usage) || die "curl failed with exit code
$?"
echo $EXPORT_RESPONSE
```

Managing multiple engines for masking

This section contains the following topic:

- [Introduction \(Managing multiple engines for masking\)](#)
- [Sync concepts](#)
- [Sync endpoints](#)
- [Key management](#)
- [Algorithm syncability](#)
- [User workflow examples](#)
- [Change log](#)

Introduction (Managing multiple engines for masking)

Your organization may have more than one masking engine, and in certain circumstances, it may want to coordinate the operation of those engines. In particular, there are two specific scenarios in which an organization could benefit from some level of interaction and orchestration between multiple masking engines.

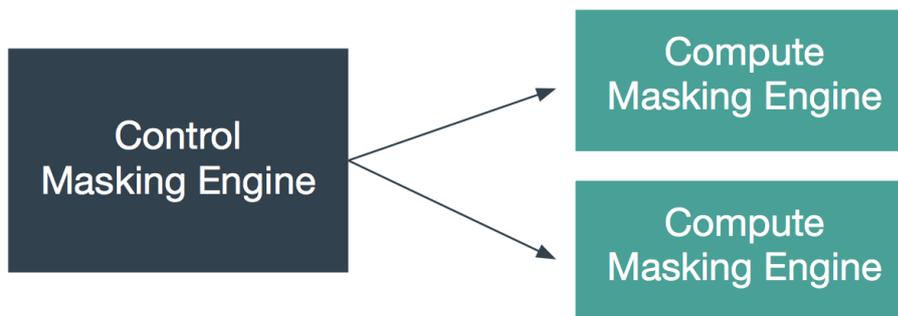
Software Development Life Cycle (SDLC)

Using an SDLC process often requires setting up multiple masking engines, each for a different part of the cycle (Development, QA, Production).



Horizontal scale

For many organizations, the size of the profiling and masking workloads requires more than one production masking engine. These masking engines can be identical in configuration or be partially equivalent depending on the organization's needs.



Best practice guide and example architectures for synchronizing

Both of these use cases require various objects to be moved between masking engines, such as Connectors, Rule Sets, and more. Engine synchronization provides a general and flexible way to move the objects necessary to run an identical job on another engine. The following sections describe how to use the Masking APIs to accomplish this.

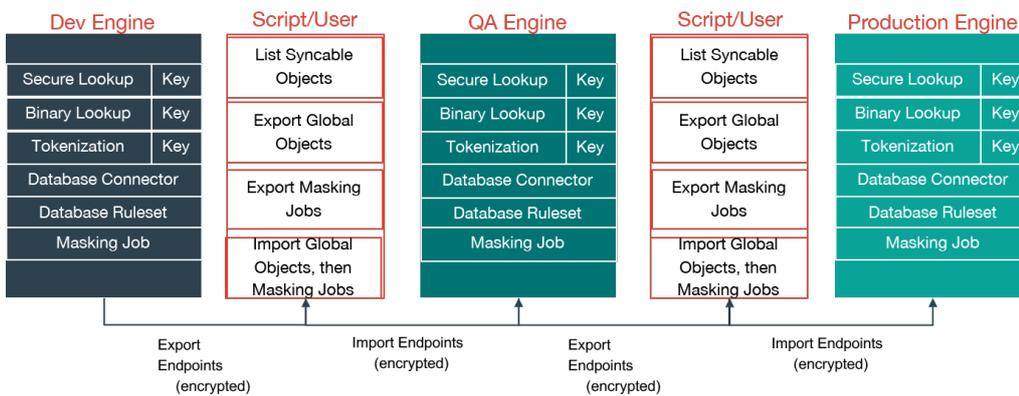
It is recommended that the syncable objects move in **only one direction**. That is, objects should be exported from one engine and imported into others but should not go in the other direction. This recommendation is primarily to simplify management of which objects exist on which engine.

For each of the scenarios above, an example architecture is described below. Note that the two architectures could be combined by having multiple production engines instead of a single one.

SDLC

The first architecture addresses the desire to author algorithms on one engine, to test and certify them on another, and finally to deploy them to a production engine. Here, algorithms are authored on the first engine, labeled “Dev Engine” in the diagram below. When the developer is satisfied, the algorithms are exported from the Dev Engine and imported to the QA Engine where they can be tested and certified. Finally, they are exported from the QA engine and imported to the production engine.

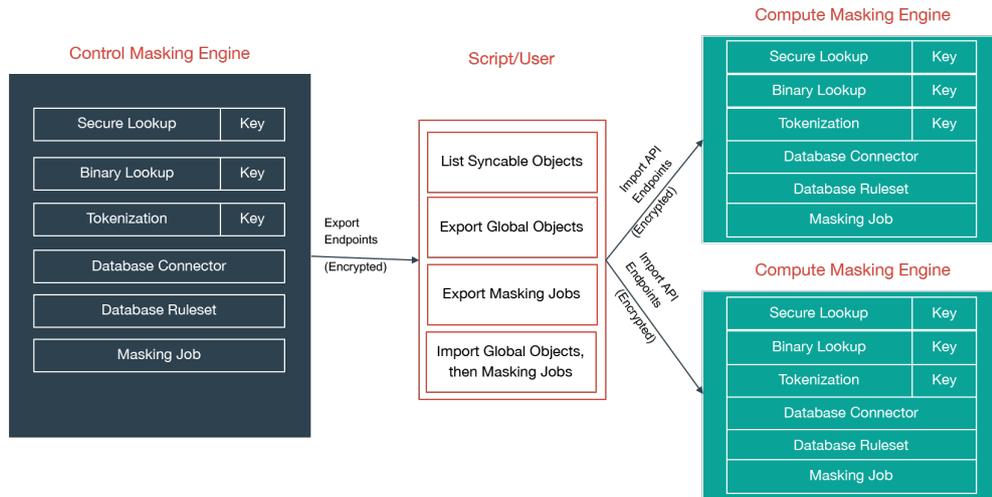
SDLC (Algorithm) Use Case



Horizontal Scale

The second architecture aims to address the problem of horizontal scale -- that is, achieving consistent masking across a large data estate by deploying multiple masking engines. In this architecture, syncable objects are authored on one engine, labeled “Control Masking Engine” in the diagram below. Those objects are then distributed to “Compute Masking Engines” using the engine synchronization APIs. The synchronized algorithms and masking jobs will produce the same masked output on all of the engines, thus enabling large data estates to be masked consistently.

Horizontal Scale Use Case



Sync concepts

Syncable object

Syncable objects are external representations of masking engine objects that can be exported from one engine and imported into another. Sync currently supports exporting a subset of algorithms (see [Algorithm Syncability](#) for details), the encryption key and all the objects necessary for a job.

 Sync does not currently support the following object(s):

- Applications

 Forward compatibility is not supported for engine sync, meaning that sync bundles from newer version engines may not import successfully into older version engines. If attempted, this could potentially result in an unexpected state or an error on the older version engine. However, backwards compatibility is supported and sync bundles from older version engines will import as expected into newer version engines, unless the sync bundle contains objects for a deprecated feature that no longer exists on the newer version engine.

Object identifiers and types

Sync uses object identifiers to name unique objects within the engine. The `/syncable-objects` endpoint provides a list of all object identifiers for a particular object type.

The following object types are currently supported:

- ALGORITHM_PLUGIN
- APPLICATION_SETTINGS
- DATABASE_CONNECTOR
- DATABASE_RULESET
- DATASET_CONNECTOR
- DATASET_FORMAT
- DATASET_RULESET
- DOMAIN
- ENVIRONMENT
- FILE_CONNECTOR
- FILE_FORMAT
- FILE_RULESET
- GLOBAL_OBJECT
- JDBC_DRIVER
- KEY
- Certain algorithms:
 - BINARYLOOKUP
 - CLEANSING
 - DATE_SHIFT
 - LOOKUP
 - MIN_MAX
 - REDACTION
 - SEGMENT
 - TOKENIZATION

- MASKING_JOB
- MOUNT_INFORMATION
- PROFILE_EXPRESSION
- PROFILE_JOB
- PROFILE_SET
- REIDENTIFICATION_JOB
- TOKENIZATION_JOB
- USER_ALGORITHM

The following lists the object types that are simply for the purpose of referencing a particular state of the exported object. These are not meant to be exported by request. The functions of these are further explained in the latter sections.

- ALGORITHM_REFERENCE
- DOMAIN_REFERENCE
- PROFILE_EXPRESSION_REFERENCE
- PROFILE_SET_REFERENCE
- SOURCE_DATABASE_CONNECTOR
- SOURCE_FILE_CONNECTOR

Dependencies



When exporting masking objects, a single export cannot contain multiple objects with the same name (e.g., two connectors with the same name).

Most objects within the Masking Engine are compositional. In order to properly capture the behavior of a syncable object, you must export its dependencies along with the object itself. Fortunately, all the necessary dependencies are exported along with the object you request; thus, it is not something you need to keep track of and worry about.

Syncable Object dependencies relationship

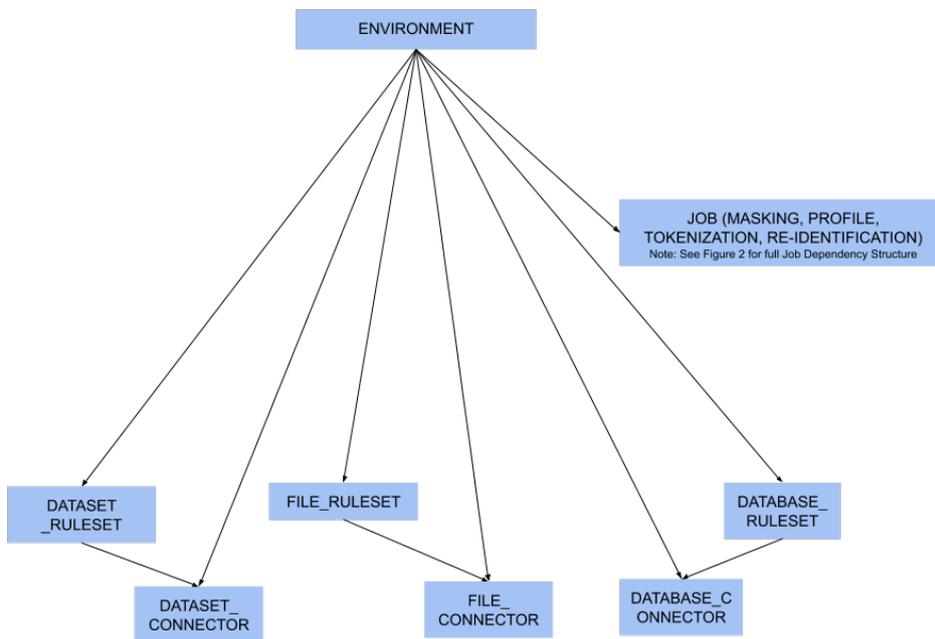


Figure 1 - environment dependencies

F While rulesets and connectors are dependencies for Jobs (see Figure 2), you may also have rulesets and connectors that are not assigned to a job. In this case, they are considered to be direct dependencies for an environment.

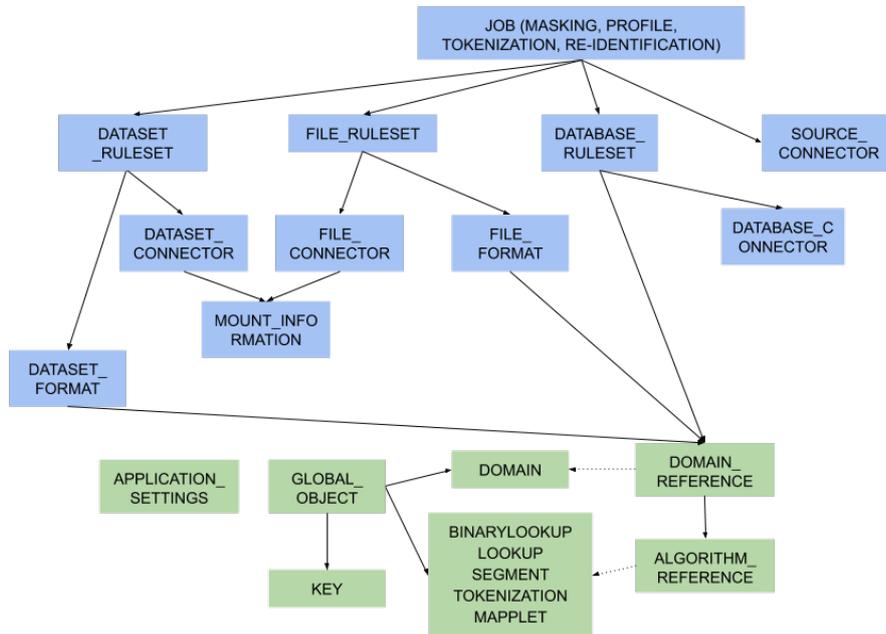


Figure 2 - object dependencies

[-] Green represents global objects (objects that are central to the entire engine), and blue represents objects that need to be a part of an environment

Object revision tracking

The revision hash is used to help you determine whether the behavior of a syncable object is the same between engines. Because objects within the Masking Engine are compositional, the behavior of an object is influenced by all of its dependencies. When a syncable object is listed or exported, the Masking Engine computes a revision_hash, which uniquely identifies the object’s behavior.

The revision_hash is a SHA1 hash that represents that object’s state, as well as the state of all objects it depends on. If two objects have the same revision hash, it is safe to assume that the behavior of the objects is the same. However, it is possible for two objects to have the same behavior but have divergent revision hashes. For example, you could have two lookup algorithms with the same name, lookup file, and key, and they do not necessarily guarantee to have the same revision hash.

[-] The revision_hash does not change when the password or the ssh key for the FILE_CONNECTOR, DATASET_CONNECTOR or DATABASE_CONNECTOR is updated. This is intentionally done because we do not export the password or the ssh key for security purposes. This allows users to update the password after import without changing the revision_hash. If a user is **overriding** a connector that already has a password set, the import **does not** reset the password and will leave the current, pre-import value.

[-] The `revision_hash` may change from version to version, and the hash comparison should be done only if both the source engine and the target engine are on the same version of the product. It is also not guaranteed to be the same between two engines at the same version if they are synced from an engine at some other version. E.g. There are three engines as follows:

- Engine A - version 5.3.2.0
- Engine B - version 5.3.3.0
- Engine C - version 5.3.3.0

If B and C are synced from A, then the `revision_hash` is not guaranteed to be the same between B and C.

Best Practice: A -> B -> C.

Export document

You can export one or more syncable objects that are listed in the `/syncable-objects` endpoint. The export document will include the set of objects that you requested for export and all of their dependencies that are required to properly import those objects into another engine.

The export document is exported as an opaque blob. Do not edit it outside of the Masking Engine.

Security

In most cases, an export document contains all the state necessary to re-create each of its objects (see [this note](#) about connector objects for one exception). In some instances, users might consider an object to be sensitive. For example, an algorithm object contains all of the information needed to produce identical algorithm results on a different engine (the algorithm's secret key, etc.). If the algorithm is being used in a production environment, then users may consider the algorithm definition and any export document containing the algorithm to be sensitive information. Therefore, export document access control, transmission, and storage should all be considered with care.

Access control

The Masking Engine only allows administrative users to make Sync API calls. When creating an administrative user account, keep in mind that the account owner will be able to access the Sync APIs to export and import objects. For this and other reasons, administrative accounts should only be created for trusted individuals.

Non-administrative accounts are not allowed to use the Sync APIs.

Transmission security

An export document containing a sensitive object should only be transmitted over a secure channel. This applies to situations where the Masking Engine is one of the transmission endpoints and when it is not. For example, when uploading (downloading) an export document to (from) the Masking Engine, the Sync API calls, like all Masking API calls, should be performed over HTTPS. Similarly, if an export document is transferred from a user's laptop to a server, the export document should be transmitted securely.

Storage security

An export document containing a sensitive object should be encrypted before it is stored persistently. Users are free to apply an encryption mechanism of their choosing to an export document. As a convenience, you can request that the export document be encrypted by the Masking Engine using a passphrase. The Masking Engine will encrypt the

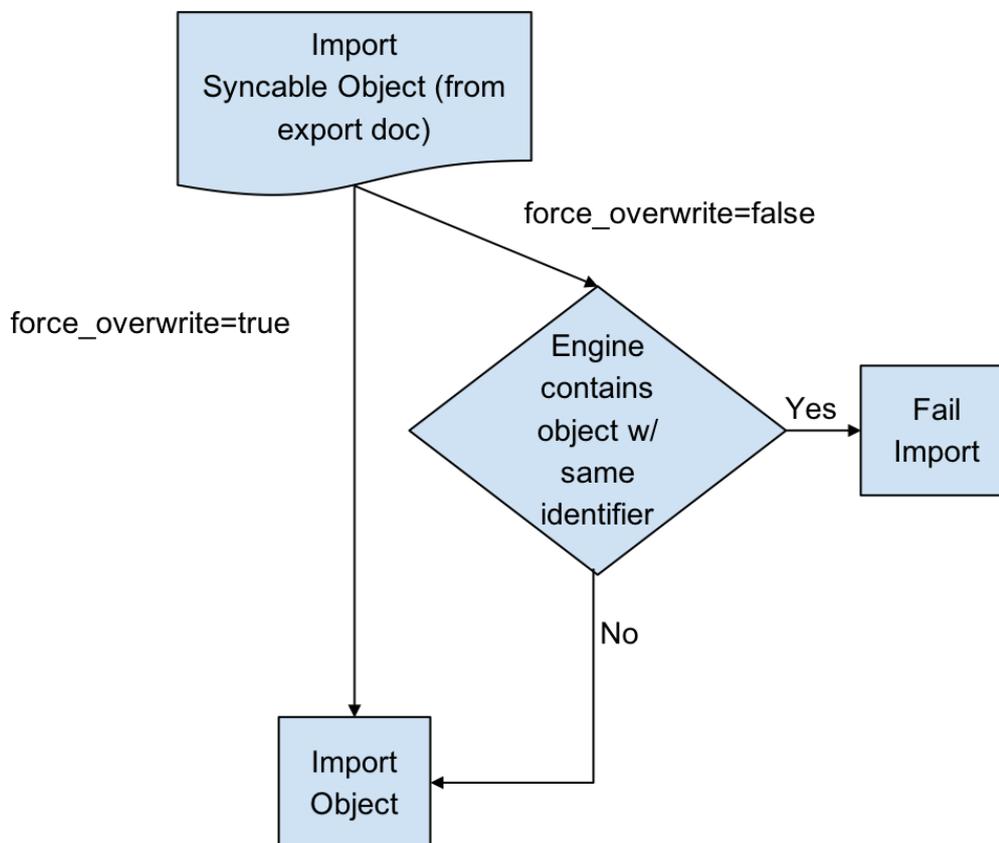
export document with 3DES using SHA1 (PBESWithSHA1AndDESede). Once the document is encrypted with the passphrase, the engine forgets the passphrase. You will need to provide the same passphrase during import to decrypt the document.

Digital signature

In order to detect accidental or malicious modification of the export document, each document is digitally signed. If the export document does not match its expected digital signature, a Masking Engine will not import the document.

Overwrite

When an object to be imported has the same name as a currently existing object, importing it will cause the other object to be changed. Since this might not be intended, we offer a flag called `force_overwrite`. If `force_overwrite` is set to false and doing the import will change an existing object on the masking engine, we fail the import. This workflow is shown below.



Attempting to import identical objects

The Masking Engine checks for the existence of the same object contents during the import of an object. If it is determined that the engine and the document being imported contain the same content, a result of SUCCESS will be returned without repeating the work of a full import. For example, importing an entire ruleset with hundreds of thousands of tables can be quite time consuming, and this should not be repeated if the same object already exists. If the object content matches and we skip the full import we note this in the application log.

Below is an example of a log statement when an identical database connector was imported:

```
2017-07-19 10:17:06,075 [http-nio-exec-4] INFO
c.d.s.marshalls.SyncableMarshaller - Skipping import process for
{
  "objectType": "DATABASE_CONNECTOR",
  "id": {
    "@type": "type.googleapis.com/IntegerIdentifier",
    "id": 1
  }
}, due to no discrepancy between the existing and importing object
```

Depending on the object type, some define an object by a String (name) and some by an Integer (object id). Objects that can have the same name in multiple environments, such as connectors, rulesets, and masking jobs, are exported based on a unique id associated with them. Global objects, which do not have overlapping names, are exported and identified based on their names. Something to note here is that objects exported based on their ids will overwrite the object with the *same name* rather than the same id. This means that for all importing objects, we define the identity of an object to be based on the name in the same environment.

For example, if I export a database connector named `testConnector` with the following export object metadata:

```
{
  "objectIdentifier": {
    "id": 5
  },
  "objectType": "DATABASE_CONNECTOR",
  "revisionHash": "68eaffef400e426520a5fcbb683419db3be53317"
}
```

And then I import this object into some engine's environment with the following list of connectors:

id	connector name	more information
1	testConnector	...
5	otherConnector	...

`testConnector` of id 1 will be overwritten, instead of `otherConnector`.

Overwrite of the encryption key

The global encryption key is somewhat special in that it always exists. Specifying `force_overwrite=false` will always fail to import the encryption key unless the encryption key has been previously synchronized using `force_overwrite=true`.

Specifying `force_overwrite=true` will always overwrite the engine's encryption key with the contents of the encryption key in the export document.

Error handling

Export documents often have multiple objects to be imported at once. For example, when exporting a database ruleset, you will export both the database ruleset and the database connector since a ruleset depends on a connector.

The engine will import one object at a time, where the dependencies are imported first. If there is an error importing an object, the import process will abort and all objects that have successfully been imported during this request will get rolled back. For example, say you are importing objects A, B, and C. Import successfully imports A. During the import of B, the engine encounters an error. The import of A will roll back, and the import of C will never execute. This will leave the engine in a state identical to the one it was in prior to the failed import.

Concurrent sync operations

To prevent race conditions with concurrent imports and jobs running, we currently do not allow concurrent import operations. We also do not allow imports while masking jobs or exports are running. It is best to do imports when a machine is not running jobs or other exports in order to guarantee that the final state of each of those operations is as expected. If they are done at the same time, the operations will fail with relevant error messages.

Global objects

GLOBAL_OBJECT is a syncable object type that is a collection of all syncable algorithms, ALGORITHM_PLUGIN(s), DOMAIN(s), JDBC_DRIVER(s), PROFILE_SET(s), PROFILE_EXPRESSION(s) and KEY (global key). This represents objects in the Masking Engine that are available across all environments, and are not a part of any specific environment. When a user requests to export GLOBAL_OBJECT, every syncable algorithm, profile set, profile expression and domain on the engine will be exported as the bundle. If a DOMAIN, PROFILE_SET, or PROFILE_EXPRESSION has a dependency on a non-syncable algorithm, such as Mapping, it will not be exported.

This separation was added because global objects 1) containing large lookup files are projected to be time-consuming and 2) are expected to be synchronized much less frequently than any masking job-related metadata. Examples on how to use it will be available in the [Example User Workflow section](#).

References objects

As mentioned in the *Global Objects* section, we expect the users to synchronize global objects and masking jobs at different frequencies. To avoid any unnecessary export of large algorithms, any objects (MASKING_JOB, PROFILE_JOB, DATABASE_RULESET, FILE_FORMAT and FILE_RULESET) that have dependencies on algorithms will export just the references to the objects by default. This way we check whether the necessary dependency exists on the importing engine by comparing the references; if not, we fail the import execution with an appropriate message. Domains, profile sets, and profile expressions are the exception to this. Exporting any of these objects will also export the full algorithm.

On-the-fly masking jobs

By definition, On-The-Fly (OTF) masking jobs work with a source environment/connector and a target environment/connector, masking the data from the source connector into that of the target connector. With masking jobs, a target *environment_id* is always required to specify which environment to import the job and its target connector. In addition to the target *environment_id*, OTF masking jobs require the specification of a *source_environment_id* into which to import the source connector. The source connector is copied into the specified source environment (*source_environment_id*), and is represented by the SOURCE_DATABASE_CONNECTOR or

SOURCE_FILE_CONNECTOR for database and file masking jobs respectively in the export document. These source connectors are virtually identical to their DATABASE_CONNECTOR and FILE_CONNECTOR counterparts, but are

represented differently in the OTF jobs to distinguish them from the target connector (i.e., DATABASE_CONNECTOR or FILE_CONNECTOR).

Circular dependencies

It is possible to have a set of objects that end up depending on each other. This would be the case if a PROFILE_SET depended on a PROFILE_EXPRESSION that depended on a DOMAIN that depended on a REDACTION algorithm that depended on the original PROFILE_SET. The masking application will detect such scenarios on export and refuse to export such configurations.

This can be worked around by creating a second PROFILE_SET that contains PROFILE_EXPRESSIONS that do not depend on a DOMAIN that depends on a REDACTION algorithm. Simply ensure that the regular expressions are the same in the newly created PROFILE_EXPRESSIONS and assign the REDACTION algorithm to the new PROFILE_SET instead. The REDACTION algorithm will function the same but the dependency loop will have been broken.

Sync endpoints

- When exporting masking objects, a single export cannot contain multiple objects with the same name (e.g., two connectors with the same name).

GET /Syncable-objects

```
GET /syncable-objects[?object_type=<type>]
```

This endpoint lists all objects in an engine that are syncable and can be exported. Any object which can be exported can be imported into another engine. The endpoint takes an optional parameter to filter by a specific object type. Each object is listed with its revision_hash. Note that if a syncable object depends on a non-syncable object (e.g. DOMAIN using a mapping algorithm), it will say so in the “revisionHash” attribute, and will not be exportable.

Example CURL command using the object_type parameter to only retrieve the list of LOOKUP algorithm objects:

```
curl -X GET
--header 'Accept: application/json'
--header 'Authorization: 21c45f0e-82f4-4b04-9072-b49072986231'
'http://masking-engine.com/masking/api/syncable-objects?object_type=LOOKUP'
```

POST /export

This endpoint allows you to export one or more objects in batch fashion. The result of the export is an export document and a set of metadata that describes what was exported. You are expected to specify which objects to export by copying their object identifiers from the /syncable-objects endpoint.

- The Sync POST /export API will timeout after 3 minutes.
- To upload objects that takes more than 3 minutes of time in uploading, use the export-async API.

The endpoint has a single optional header, *passphrase*. If you provide the passphrase, the export document will be encrypted using it.

Example CURL command using the optional passphrase header:

```
curl -X POST
--header 'Content-Type: application/json'
--header 'Accept: application/json'
--header 'Authorization: 21c45f0e-82f4-4b04-9072-b49072986231'
--header 'passphrase: my example passphrase'
-d '[
{
"objectIdentifier": {"id": 1},
"objectType": "MASKING_JOB",
"revisionHash": "asdfjkl12jijfdsaklfj21ojasdk"
}
```

```

}
]'
```

```
'http://masking-engine.com/masking/api/export'
```

POST /export-async

This endpoint does exactly the same thing as /export, but the execution is done asynchronously. The response returns an async task in the form of this:

```

{
  "asyncTaskId": 66,
  "operation": "EXPORT",
  "reference": "EXPORT-ZXhwb3J0X2RvY3VtZW50XzJjcm1EV09yLmpzb24=",
  "status": "RUNNING",
  "startTime": "2018-04-13T17:49:55.354+0000",
  "cancellable": false
}
```

Example CURL command:

```

curl -s -X POST
--header 'Content-Type: application/json'
--header 'Accept: application/json'
--header 'Authorization: 21c45f0e-82f4-4b04-9072-b49072986231'
-d "[
{
  "objectIdentifier": {"id": 1},
  "objectType": "MASKING_JOB",
  "revisionHash": "asdfjkl12jijfdsaklfj21ojasdk"
}
]"
"http://masking-engine.com/masking/api/export-async"
```

The *reference* is used to retrieve the export document of completed async export tasks from the /file-downloads endpoint. The downloaded file from this reference should look exactly the same as the response from /export.

Example CURL command:

```

curl -s -X GET
--header 'Accept: application/octet-stream'
--header 'Authorization: 21c45f0e-82f4-4b04-9072-b49072986231'
-o "<OUTPUT_FILE_PATH>" "http://masking-engine.com/masking/api/file-downloads/EXPORT-ZXhwb3J0X2RvY3VtZW50XzJjcm1EV09yLmpzb24="
```

Error handling

If an error occurs while exporting one or more elements in the export document, the entire export will abort.

POST /import

```
POST /import?force_overwrite=<true|false>[&environment_id=<id>]
[&source_environment_id=<id>]
```

This endpoint allows you to import a document exported from another engine. The response returns a list of objects that were imported and whether the import was successful.

The endpoint has one required parameter, *force_overwrite*, two optional parameters *environment_id* and *source_environment_id*, and an optional HTTP header, *passphrase*, which if provided, will cause the engine to attempt to decrypt the document using the specified passphrase. The required *force_overwrite* parameter dictates how to deal with conflicting objects. *environment_id* is necessary for all non-global objects that need to belong in an environment. *source_environment_id* is used for On-The-Fly masking jobs.

The endpoint has a single optional header, *passphrase*. If you provide the passphrase, the import document will be decrypted using it.

⚠ Containerized Masking does not support some objects which might be exported from a Virtual Machine Masking Engine. Containerized Masking will generate an error if an import bundle contains one of these objects. To successfully import environments that contain these objects on a containerized engine, the problem objects will have to be removed on the export source engine and re-exported.

Unsupported objects include:

- Connectors using the FTP connection method
- Connectors using Kerberos credentials for DB authentication
- Connectors using IBM's custom DB2 JDBC driver
- Connectors using OAUTH2 authentication
- Engine Setup based NFS/CIFS mounts

Example CURL command using the optional passphrase header:

```
curl -X POST
--header 'Content-Type: application/json'
--header 'Accept: application/json'
--header 'Authorization: 21c45f0e-82f4-4b04-9072-b49072986231'
--header 'passphrase: my example passphrase'
-d '{
  "exportResponseMetadata": {
    "exportHost": "masking-engine.com",
    "exportDate": "Mon Aug 13 16:29:30 UTC 2018",
    "exportedObjectList": [
      {
        "objectIdentifier": {
          "algorithmName": "lookup_alg"
        },
        "objectType": "LOOKUP",
        "revisionHash": "cf84d82c21f0e9d4105d37ae7979c0848486d861"
      },
      {
        "objectIdentifier": {
          "keyId": "global"
        }
      }
    ]
  }
}
```

```

"objectType": "KEY",
"revisionHash": "1d8e9bc552d3ca1dcd218f9e197ea3955ccc29be"
}
],
},
"blob": "<OMITTED>",
"signature": "<OMITTED>", \
"publicKey": "<OMITTED>" \
}'
'http://masking-engine.com/masking/api/import?force_overwrite=true'

```

POST /import-async

```

POST /import-async?force_overwrite=<true | false>[&environment_id=<id>]
[&source_environment_id=<id>]

```

This endpoint does exactly the same thing as /import, but the execution is done asynchronously and the body is taken in as a file. The response returns an async task in the form of this:

```

{
  "asyncTaskId": 67,
  "operation": "IMPORT",
  "reference": "IMPORT-ZXhwb3J0X2RvY3VtZW50XzJjcm1EV09yLmpzb24=",
  "status": "RUNNING",
  "startTime": "2018-04-13T17:49:55.354+0000",
  "cancellable": false
}

```

The *reference* is used to retrieve the import status of completed async import tasks from the /file-downloads endpoint. The downloaded file from this reference should look exactly the same as the response from /import.

Example CURL command:

```

curl -s -X POST
--header 'Content-Type: multipart/form-data'
--header 'Accept: application/json'
--header 'Authorization: 21c45f0e-82f4-4b04-9072-b49072986231'
-F "file=@<DOWNLOADED_FILE_PATH>"
'http://masking-engine.com/masking/api/import-async?force_overwrite=true'

```

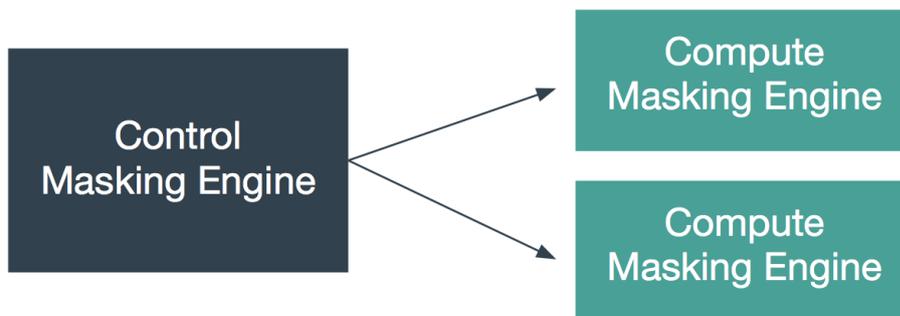
Key management

One important piece of data used by many masking algorithms is the key, which determines the masked outcome of some value. Changing the key changes the output of these algorithms. For example, if the FIRST NAME algorithm masks “Michelle” to “Rachael,” changing the key might cause it to mask “Michelle” to “Ben”. There are two types of keys that the algorithms can depend on: either 1) global key or 2) individual key.

Global key

A user with Administrator privileges can change the key by clicking the **Generate New Key** button in the **Admin** tab.

i Other actions are not allowed during the key generation process. Wait for the **Generate New Key** process to complete and a success dialogue to display in the user interface before performing additional actions on the Masking Engine (e.g., running a masking job).



Synchronizing the global Key between multiple engines

In order for algorithms to behave the same way across several engines, all of those engines must have the same global key. Changing an engine’s global key alters the behavior of all of the algorithms that depend on the global key.

You may want to change the key from time to time as a security management practice. If so, change it on all of the engines at the same time. That is, generate a new key on one engine, export that key, and import it to all of the other engines in the deployment.

Keys can be imported and exported independently of algorithms. To export the key from an engine, login to the engine through the login endpoint and then call export with the body shown below. Like all objects, you can encrypt the payload by supplying a passphrase header.

```
[{
  "objectIdentifier": {
    "keyId": "global"},
  "objectType": "KEY"
}]
```

The API will return a JSON payload containing an encoded form of the key that you can install on other engines through the import endpoint. Like all exported objects, it is encoded in an opaque blob.

Individual key

The following algorithm types have their own key that determines the masked results:

- BINARYLOOKUP
- DATE_SHIFT (only applies to DateShiftDiscrete)
- LOOKUP
- TOKENIZATION

The keys for each algorithm gets exported and imported with the algorithm itself, not separately. These individually associated keys can be randomized with an endpoint.

```
PUT http://masking-engine-A/masking/api/algorithms/{algorithmName}/randomize-key
```

Algorithm syncability

Overview

This article shows tables that specify which algorithms are syncable between Masking engines (in addition to the Masking engine key).

 Users must have admin privileges on the Masking engine to export and import algorithms.

User-defined algorithms

Type	Syncable	Notes
Lookup	Yes	NA
Binary Lookup	Yes	NA
Segmented Mapping	Yes	NA
Mapping	Configuration Specific	See Masking Algorithm Sync
Tokenization	Yes	NA
Minmax	Yes	NA
Cleansing	Yes	NA
Free Text Redaction	Yes	NA
Component	Yes	NA

Built-In algorithms

While some of the built-in algorithms are not synchronizable, mainly due to them being non-deterministic, we still can support the export of inventories that contain any built-in algorithm. We just do not guarantee consistent masking of those non-synchronizable built-in algorithms between engines.

 Syncing built-in algorithms does not actually import the files associated with them, rather, it updates their individual keys if they have them.

- = Synchronizing the global Engine key has no impact on SDK based algorithms. SDK algorithms must be synchronized as separate individual objects due to the embedded key per-algorithm object. Many out-of-the-box algorithms are transitioning to implementations based on SDK-derived plugin implementations. A user must sync the individual algorithms listed under "GET /syncable-objects" `object_type == "USER_ALGORITHM"` .

Algorithm API Name	Algorithm UI Name	Type	Syncable	Workaround
AccNoLookup	ACCOUNT SL	lookup	Yes	NA
AccountTK	ACCOUNT_TK	tokenization	Yes	NA
AddrLine2Lookup	ADDRESS LINE 2 SL	lookup	Yes	NA
AddrLookup	ADDRESS LINE SL	lookup	Yes	NA
BusinessLegalEntityLookup	BUSINESS LEGAL ENTITY SL	lookup	Yes	NA
CommentLookup	COMMENT SL	lookup	Yes	NA
CreditCard	CREDIT CARD	calculated	No	None
DateShiftDiscrete	DATE SHIFT(DISCRETE)	calculated	Yes	NA
DateShiftFixed	DATE SHIFT(FIXED)	calculated	No	Already synchronized
DateShiftVariable	DATE SHIFT(VARIABLE)	calculated	No	None
DrivingLicenseNoLookup	DR LICENSE SL	lookup	Yes	NA
DummyHospitalNameLookup	DUMMY_HOSPITAL_NAME_SL	lookup	Yes	NA
EmailLookup	EMAIL SL	lookup	Yes	NA
FirstNameLookup	FIRST NAME SL	lookup	Yes	NA
FullNMLookup	FULL_NM_SL	lookup	Yes	NA

Algorithm API Name	Algorithm UI Name	Type	Syncable	Workaround
LastNameLookup	LAST NAME SL	lookup	Yes	NA
LastCommaFirstLookup	LAST_COMMA_FIRST_SL	lookup	Yes	NA
NameTK	NAME_TK	tokenization	Yes	NA
NullValueLookup	NULL SL	lookup	Yes	NA
RandomValueLookup	RANDOM_VALUE_SL	lookup	Yes	NA
SchoolNameLookup	SCHOOL NAME SL	lookup	Yes	NA
SecureShuffle	SECURE SHUFFLE	calculated	No	None
SsnTK	SSN_TK	tokenization	Yes	NA
USCountiesLookup	US_COUNTIES_SL	lookup	Yes	NA
USCitiesLookup	USCITIES_SL	lookup	Yes	NA
USstatecodesLookup	USSTATE_CODES_SL	lookup	Yes	NA
USstatesLookup	USSTATES_SL	lookup	Yes	NA
WebURLsLookup	WEB_URLS_SL	lookup	Yes	NA
RepeatFirstDigit	ZIP+4	calculated	No	Already synchronized

Extensible algorithms

Extensible Algorithms are fully syncable between Masking Engines. There are two types of Extensible Algorithms:

1. Built-in to the Algorithm Plugin: These are synced through synchronization of the corresponding plugin.
2. Configurable Extensible Algorithms: These may be synced alone. All the dependencies for Extensible Algorithms (like other Extensible Algorithms, files, external file URLs, mount points for NFS mounted files, or algorithm plugins) are automatically synced along with it. For example: if an Extensible Algorithm is configured for use by another Extensible Algorithm, the dependent one (or containing the plugin) is automatically synced during the sync of the main one.

User workflow examples

This page provides some examples of some typical user workflows. More information on exactly how each endpoint works is available on the [Sync endpoints](#) section.

Syncing all global objects

The following steps can be used to sync all global objects from Masking Engine A to Masking Engine B. This will sync all algorithms and domains and should be done prior to syncing jobs or rulesets which might depend on them. For more information on the global object, see the [Sync concepts](#) section.

Source masking engine steps

1. Login

Login on the source Masking Engine to obtain an Authorization token value.

```
POST https://a.example.com/masking/api/login
```

HEADER

```
Content-Type: application/json
```

```
Accept: application/json
```

BODY

```
{  
  "username": "user123",  
  "password": "pw123"  
}
```

CURL example:

```
curl -X POST --cacert /path/to/cert --header 'Content-Type: application/json' --  
header 'Accept: application/json' -d '{ "username": "user123", "password": "pw123" }'  
'https://a.example.com/masking/api/login'
```

Expected Result:

```
{  
  "Authorization": "dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a"  
}
```

2. Get the identifier

Call `GET /syncable-objects` to obtain the GLOBAL_OBJECT's information.

```
GET https://a.example.com/masking/api/syncable-objects?object_type=GLOBAL_OBJECT
```

HEADER

```
Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a (value from the /login response)
Accept: application/json
```

CURL example:

```
curl -X GET --cacert /path/to/cert --header 'Accept: application/json' --header
'Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a' 'https://a.example.com/masking/
api/syncable-objects?object_type=GLOBAL_OBJECT'
```

Expected Result:

```
{
  "_pageInfo": {
    "numberOnPage": 1,
    "total": 1
  },
  "responseList": [
    {
      "objectIdentifier": {
        "id": "global"
      },
      "objectType": "GLOBAL_OBJECT",
      "revisionHash": "8d5236bb029c2176aa568b930786b63253e4f9e4"
    }
  ]
}
```

3. Export the object

Call `POST /export-async` to asynchronously export the GLOBAL_OBJECT and use the passphrase header to encrypt the export.

 The optional passphrase header cannot be specified using the interactive [API Client tool](#). An example of how to specify this header using a cURL command is shown below.

```
POST https://a.example.com/masking/api/export-async
```

HEADER

```
Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a
Content-Type: application/json
Accept: application/json
passphrase: my example passphrase
```

BODY

```
[
  {
    "objectIdentifier": {
      "id": "global"
    },
  },
]
```

```

    "objectType": "GLOBAL_OBJECT"
  }
]

```

```

curl -X POST --cacert /path/to/cert --header 'Content-Type: application/json' --header 'Accept: application/json' --header 'Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a' --header 'passphrase: my example passphrase' -d '[{"objectIdentifier":{"id":"global"},"objectType":"GLOBAL_OBJECT"}]' 'https://a.example.com/masking/api/export-async'

```

Expected Result:

```

{
  "asyncTaskId": 2,
  "operation": "EXPORT",
  "reference": "EXPORT-ZXhwb3J0X2RvY3VtZW50Xzk0Wjlvva3JDLmpzb24=",
  "status": "RUNNING",
  "startTime": "2018-06-15T20:36:35.483+0000",
  "cancellable": false
}

```

4. Download the export document

Use the reference above to download the export document via the /file-download endpoint.

```

GET https://a.example.com/masking/api/file-downloads/EXPORT-ZXhwb3J0X2RvY3VtZW50Xzk0Wjlvva3JDLmpzb24=

```

```

HEADER
Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a
Accept: application/octet-stream

```

CURL example:

```

curl -X GET --cacert /path/to/cert --header 'Accept: application/json' --header 'Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a' 'https://a.example.com/masking/api/file-downloads/EXPORT-ZXhwb3J0X2RvY3VtZW50Xzk0Wjlvva3JDLmpzb24='

```

Expected Result: An export document that will look like this.

```

{
  "exportResponseMetadata": {
    "exportHost": "a.example.com",
    "exportDate": "Fri Jun 15 20:16:20 UTC 2018",
    "requestedObjectList": [
      {
        "objectIdentifier": {

```

```

      "id": "global"
    },
    "objectType": "GLOBAL_OBJECT",
    "revisionHash": "579850b1c88baf74cee6bad61d81e2aa3dcc206c"
  }
],
"exportedObjectList": [
  {
    "objectIdentifier": {
      "id": "DRIVING_LC"
    },
    "objectType": "DOMAIN",
    "revisionHash": "9ee90782488d14d369f9595dad7f593c961e785f"
  },
  {
    "objectIdentifier": {
      "algorithmName": "DrivingLicenseNoLookup"
    },
    "objectType": "LOOKUP",
    "revisionHash": "e08ac9bfd4ed9f64d486cb47cdc07deb30ccc20f"
  },
  ...
]
},
"blob":
"RAAAAOkZmZhNWIxNjktODMwMC00N2FLLWJjZmMtNjVhNDUzYWl3OTBjEhgyMDE4LTA2LTE1VDIwOjE2OjIw
LjY2MFogBSgBFwIAAAOkZmZhNWIxNjktODMwMC00N2FLLWJjZmMtNjVhNDUzYWl3OTBjEu4DCi8IFBIrCiV0e
XBllmdvb2dsZWFWaXMuY29tL0ludGVnZXJJZGVudGlmawVvYegIARivCA4SKwoIdHlwZS5nb29nbGVhcGlzLm
...",
"signature": "MCwCFAWGF/97wb+oYuSQizj8U12n7jpQAhQKGCa0J4U8XyDA0EhMUWkzZXHrpw==",
"publicKey": "MIHxMIGoBgcqhkJ00AQBMIgcAkEA/
KaCzo4Syrom78z3EQ5SbbB4sF7ey80etKII864WF64B81uRPH5t9jQTxeEu0ImbzRMqzVDZkVG9xD7nN1kuFw
IVAJYu3cw2nLq0uyY05rahJtk0bjjFAkBnhHGyepz0TukaScUUfbGpq.."
}

```

5. Cleanup

When the export document is no longer needed, use the `/export-async` endpoint to cleanup the exported documents.

```
DELETE https://a.example.com/masking/api/export-async
```

HEADER

```
Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a
Accept: application/json
```

CURL example:

```
curl -X DELETE --header 'Accept: application/json' --header 'Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a' 'https://a.example.com/masking/api/export-async'
```

Expected Result: no content

Destination Masking Engine steps

1. Login

Login on the destination Masking Engine to obtain an Authorization token value (see example above).

2. Import the object

On Masking Engine B, use the import-async endpoint to import the document downloaded from engine A.

```
POST https://b.example.com/masking/api/import-async?force_overwrite=true
```

HEADER

```
Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a
Content-Type: multipart/form-data
Accept: application/json
passphrase: my example passphrase
```

CURL example:

```
curl -X POST --cacert /path/to/cert --header 'Content-Type: multipart/form-data' --header 'Accept: application/json' --header 'Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a' --header 'passphrase: my example passphrase' -F file=@export.json 'https://b.example.com/masking/api/import-async?force_overwrite=true'
```

Expected Result:

```
{
  "asyncTaskId": 1,
  "operation": "IMPORT",
  "reference": "IMPORT-aW1wb3J0X2RvY3VtZW50X2lZQVFKWEFsLmpzb24=",
  "status": "WAITING",
  "cancellable": false
}
```

3. Verify status

On Masking Engine B, call the /file-downloads endpoint using the reference from the returned Async Task response to retrieve the completed import status.

```
GET https://b.example.com/masking/api/file-downloads/IMPORT-aW1wb3J0X2RvY3VtZW50X2lZQVFKWEFsLmpzb24=
```

HEADER

```
Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a
Accept: application/octet-stream
```

CURL example:

```
curl -X GET --cacert /path/to/cert --header 'Accept: application/octet-stream' --header 'Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a' 'https://b.example.com/masking/api/file-downloads/IMPORT-aW1wb3J0X2RvY3VtZW50X2lZQVFKWEFsLmpzb24='
```

Expected Result:

An import status document that reports the success or failure of each object imported.

```
[
  {
    "objectIdentifier": {
      "id": 7
    },
    "importedObjectIdentifier": {
      "id": 7
    },
    "objectType": "PROFILE_EXPRESSION",
    "importStatus": "SUCCESS"
  },
  {
    "objectIdentifier": {
      "id": "CERTIFICATE_NO"
    },
    "importedObjectIdentifier": {
      "id": "CERTIFICATE_NO"
    },
    "objectType": "DOMAIN",
    "importStatus": "SUCCESS"
  },
  ...
]
```

4. Cleanup

Once the status is no longer needed, use the /import-async endpoint to cleanup the exported documents.

```
DELETE https://b.example.com/masking/api/import-async
```

```
HEADER
Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a
Accept: application/json
```

CURL example:

```
curl -X DELETE --cacert /path/to/cert --header 'Accept: application/json' --header 'Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a' 'https://b.example.com/masking/api/import-async'
```

Expected Result: no content

Syncing a masking job

The following steps provide an example of how to export a Masking Job from Masking Engine A to Masking Engine B using the synchronous endpoints of /export and /import. This presumes that all of the global objects such as algorithms and domains that the masking job relies on have already been synced. This can also be done via the asynchronous endpoint with the same workflow as above.

1. Export the job

Before this step, the /login and /syncable-objects endpoints should have been called to obtain the authorization token and job identifier respectively. Then use the /export endpoint to obtain an export document with the desired MASKING_JOB. In this example, the optional passphrase is used to encrypt the export document.

 To sync a profile job, swap out the objectType for "PROFILE_JOB" and provide the id of the profile job to sync. Profile jobs are syncable starting in version 5.3.2.0.

```
POST http://a.example.com/masking/api/export

HEADER
Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a
Content-Type: application/json
Accept: application/json
passphrase: password to encrypt the export document

BODY
[
  {
    "objectIdentifier": {
      "id": 4
    },
    "objectType": "MASKING_JOB"
  }
]
```

CURL example:

```
curl -X POST --cacert /path/to/cert --header 'Content-Type: application/json' --header 'Accept: application/json' --header 'Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a' --header 'passphrase: my example passphrase' -d '[ { "objectIdentifier": { "id": 4 }, "objectType": "MASKING_JOB" } ]' 'https://a.example.com/masking/api/export'
```

Expected Result:

```
{
  "exportResponseMetadata": {
    "exportHost": "a.example.com",
    "exportDate": "Fri Jun 15 20:16:20 UTC 2018",
    "requestedObjectList": [
```

```

    {
      "objectIdentifier": {
        "id": 1
      },
      "objectType": "MASKING_JOB",
      "revisionHash": "579850b1c88baf74cee6bad61d81e2aa3dcc206c"
    }
  ],
  "exportedObjectList": [
    {
      "objectIdentifier": {
        "id": 1
      },
      "objectType": "DATABASE_RULESET",
      "revisionHash": "bf63b401129cbc84f90eeb708281e98121f5a829"
    },
    {
      "objectIdentifier": {
        "id": "FIRST_NAME"
      },
      "objectType": "DOMAIN_REFERENCE",
      "revisionHash": "e6a52079843afd2625f20237fd50f56254c7e630"
    },
    {
      "objectIdentifier": {
        "id": 1
      },
      "objectType": "MASKING_JOB",
      "revisionHash": "579850b1c88baf74cee6bad61d81e2aa3dcc206c"
    },
    {
      "objectIdentifier": {
        "id": 1
      },
      "objectType": "DATABASE_CONNECTOR",
      "revisionHash": "6455f39dfa354a54bdf4ef69d6511a6c2bb19db3"
    },
    {
      "objectIdentifier": {
        "algorithmName": "FirstNameLookup"
      },
      "objectType": "ALGORITHM_REFERENCE",
      "revisionHash": "13b0a51a7e3904f52526c442419c54b39033dca3"
    }
  ]
},
"blob":
"RAAAAaokZmZhNWIxNjktODMwMC00N2FLLWJjZmMtNjVhNDUzYWI3OTBjEhgyMDE4LTA2LTE1VDIwOjE2OjIw
LjY2MFogBSgBFwIAAAokZmZhNWIxNjktODMwMC00N2FLLWJjZmMtNjVhNDUzYWI3OTBjEu4DCi8IFBIrCiV0e
XB\lmdvb2dsZWFWaXMuY29tL0\udGVnZXJJZGVudGlnaWVyEgIIARivCA4SKwo\ldHlwZS5nb29nbGVhcGlzLm
...",
"signature": "McwCFAWgf/97wb+oYuSQizj8U12n7jppQAhQKGCa0J4U8XyDA0EhMUWkzZXHrpw==",

```

```
"publicKey": "MIHxMIGoBgcqhkJ00AQBMIgCAkEA/
KaCzo4Syrom78z3EQ5SbbB4sF7ey80etKII864WF64B81uRpH5t9jQTxeEu0ImbzRMqzVDZkVG9xD7nN1kuFw
IVAJYu3cw2nLq0uyY05rahJtk0bjjFAkBnhHGyepz0TukaScUUfbGpq.."
}
```

 The requestedObjectList returns the list of objects you've requested in the export, and the exportedObjectList returns a list of all objects that were exported. This will include both the requested ones and their dependencies.

2. Import the job

On Masking Engine B, import the masking job. You will need to provide an environment for it to import into.

```
POST http://b.example.com/masking/api/import?force_overwrite=false&environment_id=1
```

HEADER

```
Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a
Content-Type: application/json
Accept: application/json
passphrase: password to encrypt the export document
```

PARAMETER

force_overwrite and environment_id. See the details in the Masking API Call Concepts section [for](#) more details .

BODY

(Whatever gets returned from export)

CURL example:

```
curl -X POST --cacert /path/to/cert --header 'Content-Type: application/json' --
header 'Accept: application/json' --header 'Authorization: dc2cff8b-
e20d-4e28-8b7e-5d7c4aad0e2a' --header 'passphrase: my example passphrase' -d @/path/
to/export.json 'https://b.example.com/masking/api/import?
force_overwrite=false&environment_id=1'
```

Expected Result:

```
[
  {
    "objectIdentifier": {
      "id": 3033
    },
    "importedObjectIdentifier": {
      "id": 1
    },
    "objectType": "DATABASE_CONNECTOR",
    "importStatus": "SUCCESS"
  }
]
```

```

    },
    {
      "objectIdentifier": {
        "id": 5421
      },
      "importedObjectIdentifier": {
        "id": 1
      },
      "objectType": "DATABASE_RULESET",
      "importStatus": "SUCCESS"
    }
    ...
  ]

```

Syncing an environment

Syncing an environment differs from syncing other objects in that we don't sync any of the environment's metadata, only its dependencies (jobs, connectors and rulesets). You can think of syncing an environment as an easy way to sync a large group of objects in the environment, without having to sync them one at a time. As such, the environment's revisionHash is not important.

1. Export the environment

Before this step, the `/login` and `/syncable-objects` endpoints should have been called to obtain the authorization token and environment identifier respectively. Then use the `/export` endpoint to obtain an export document with the desired ENVIRONMENT. In this example, the optional passphrase is used to encrypt the export document.

```

POST http://a.example.com/masking/api/export

HEADER
Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a
Content-Type: application/json
Accept: application/json
passphrase: password to encrypt the export document

BODY
[
  {
    "objectIdentifier": {
      "id": 3
    },
    "objectType": "ENVIRONMENT"
  }
]

```

CURL example:

```
curl -X POST --cacert /path/to/cert --header 'Content-Type: application/json' --
header 'Accept: application/json' --header 'Authorization: dc2cff8b-
e20d-4e28-8b7e-5d7c4aad0e2a' --header 'passphrase: my example passphrase' -d
'[ { "objectIdentifier": { "id": 3 }, "objectType": "ENVIRONMENT" } ]' 'https://
a.example.com/masking/api/export'
```

Expected Result:

```
{
  "exportResponseMetadata": {
    "exportHost": "a.example.com",
    "exportDate": "Tue Apr 21 21:57:32 UTC 2020",
    "requestedObjectList": [
      {
        "objectIdentifier": {
          "id": 3
        },
        "objectType": "ENVIRONMENT",
        "revisionHash": "c2f2f4bd8a043c32d0977cff8f915d64f1aaf518"
      }
    ],
    "exportedObjectList": [
      {
        "objectIdentifier": {
          "id": 4
        },
        "objectType": "DATASET_CONNECTOR",
        "revisionHash": "db7bc78d098f3df47199fc00c2ba83dee5a52a34"
      },
      {
        "objectIdentifier": {
          "id": 3
        },
        "objectType": "ENVIRONMENT",
        "revisionHash": "c2f2f4bd8a043c32d0977cff8f915d64f1aaf518"
      },
      {
        "objectIdentifier": {
          "id": 4
        },
        "objectType": "MASKING_JOB",
        "revisionHash": "2497260ee897303fc317b9268486c5e36663dad0"
      },
      {
        "objectIdentifier": {
          "id": 4
        },
        "objectType": "DATASET_RULESET",
        "revisionHash": "cb864b0f3f208c4ea5273389055d335d8d57028c"
      },
      {
        "objectIdentifier": {
```

```

      "id": 1
    },
    "objectType": "DATASET_FORMAT",
    "revisionHash": "0513a494c736d7f8993dee4720f200c0aa3bd749"
  }
]
},
"blob": "RAAAAokZDg5Zjg5NWQtYzJjMi00ZjkyLWlXNjEtMTA0NDRjZDk5YWlxEhgyMDI...",
"signature": "MCwCF9wqsdqMG/x7q+knwd4LLhwc4h+AhR9YF5rQZyp5YLQf8e7rI39kjkyUQ==",
"publicKey": "MIHwMIGoBgcqhkJ00AQBMIGcAKEA/KaCzo4Syrom78z3EQ5SbbB4sF7ey8..."
}

```

- [-] The requestedObjectList returns the list of objects you've requested in the export, and the exportedObjectList returns a list of all objects that were exported. This will include both the requested ones and their dependencies.

2. Create a new environment on the target engine

Since we do not import the environment metadata (such as name or type) we must first create an environment on the target which we wish to import our data into. At this step we would also need to create a source environment if we are importing any On-The-Fly jobs.

- [-] All source connectors will end up being imported into the source environment that we specify. If you wish for these to be in separate environments they must then be manually managed after import.

3. Import the environment into the newly created environment

On Masking Engine B, import the environment. You will need to provide an environment for it to import into.

```
POST http://b.example.com/masking/api/import?force_overwrite=false&environment_id=1
```

HEADER

```
Authorization: dc2cff8b-e20d-4e28-8b7e-5d7c4aad0e2a
```

```
Content-Type: application/json
```

```
Accept: application/json
```

```
passphrase: password to encrypt the export document
```

PARAMETER

force_overwrite and environment_id. See the details in the Masking API Call Concepts section [for](#) more details .

BODY

(Whatever gets returned from export)

CURL example:

```
curl -X POST --cacert /path/to/cert --header 'Content-Type: application/json' --
header 'Accept: application/json' --header 'Authorization: dc2cff8b-
e20d-4e28-8b7e-5d7c4aad0e2a' --header 'passphrase: my example passphrase' -d @/path/
to/export.json 'https://b.example.com/masking/api/import?
force_overwrite=false&environment_id=1'
```

Expected Result:

```
[
  {
    "objectIdentifier": {
      "id": 4
    },
    "importedObjectIdentifier": {
      "id": 5
    },
    "objectType": "DATASET_CONNECTOR",
    "importStatus": "SUCCESS"
  },
  {
    "objectIdentifier": {
      "id": 3
    },
    "importedObjectIdentifier": {
      "id": 1
    },
    "objectType": "ENVIRONMENT",
    "importStatus": "SUCCESS"
  },
  {
    "objectIdentifier": {
      "id": 1
    },
    "importedObjectIdentifier": {
      "id": 1
    },
    "objectType": "DATASET_FORMAT",
    "importStatus": "SUCCESS"
  },
  {
    "objectIdentifier": {
      "id": 4
    },
    "importedObjectIdentifier": {
      "id": 5
    },
    "objectType": "DATASET_RULESET",
    "importStatus": "SUCCESS"
  },
  {
    "objectIdentifier": {
      "id": 4
    }
  }
]
```

```
  },  
  "importedObjectIdentifier": {  
    "id": 5  
  },  
  "objectType": "MASKING_JOB",  
  "importStatus": "SUCCESS"  
}  
]
```

Change log

Changes in 6.0

New syncable objects

We added the following new syncable objects in 6.0. Refer to the main documentation for more information on what they are, and how to use them.

- 6.0.0.0 Release
 - MOUNT_INFORMATION
- 6.0.1.0 Release
 - JDBC_DRIVER
 - REIDENTIFICATION_JOB
 - TOKENIZATION_JOB
- 6.0.2.0 Release
 - DATASET_CONNECTOR
 - DATASET_FORMAT
 - DATASET_RULESET
 - ENVIRONMENT
- 6.0.3.0 Release
 - ALGORITHM_PLUGIN
 - USER_ALGORITHM

Changes in 5.3

New syncable objects

We added the following new syncable objects in 5.3. Refer to the main documentation for more information on what they are, and how to use them.

- 5.3.0.0 Release
 - DATABASE_RULESET
 - DATE_SHIFT
 - DOMAIN
 - FILE_CONNECTOR
 - FILE_FORMAT
 - FILE_RULESET
 - GLOBAL_OBJECT
 - MASKING_JOB
- 5.3.3.0 Release
 - PROFILE_EXPRESSION
 - PROFILE_JOB
 - PROFILE_SET

We also added the following new syncable algorithms in 5.3.

- 5.3.2.0 Release
 - CLEANSING
 - MIN_MAX
- 5.3.3.0 Release
 - REDACTION

Key per algorithm

In pre-5.3, a global key for the engine was used by all algorithms that required a seed to determine the outcome of masked values. This included algorithms such as Lookup and Binary Lookup. Thus, in 5.2, exporting a Lookup Algorithm would automatically export the global encryption key as a dependency. In this release, we allow each algorithm to have its own independent key, exported as a part of the algorithm. Refer to the [Key Management](#) section for more detail.

Changed model of import status reporting

In 5.2, the import status looked like this: some browsers enable drag-n-drop only when dataTransfer has data

```
{
  "objectIdentifier": {
    "keyId": "global"
  },
  "objectType": "KEY",
  "importStatus": "SUCCESS"
}
```

Starting in 5.3.0, the import status of an object has extended to include the id or name it has imported into to reduce any confusion introduced with IntegerIdentifiers. For more information on the reason for this change, refer to Logic Behind Overwrite of IntegerIdentifier and StringIdentifier. For examples on what it now looks like, refer to the [Example User Workflow](#) section.

Changed granularity of transactions for import

Starting in 5.3, an import of however many objects is performed as an atomic execution rather than using per-object atomicity. This means that the execution will either succeed at importing all objects or fail and import none at all. Refer to the Error Handling of Import logic flow diagram for more information.

Filter for /syncable-objects

Now that we have a large list of syncable objects, we have added a new feature for filtering based on the object type. Refer to the [Endpoints](#) page and the [Example User Workflow](#) section for more information.

Async endpoints

Exporting a large MASKING_JOB with many dependencies can potentially take a long time. So we have decided to provide a new endpoint that exports and imports the objects asynchronously. Refer to the [Endpoints](#) section in the main documentation and the [Example User Workflow](#) page for more information.